

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 25, 2001, 11:43:24 : Search time 12.34 Seconds
(without alignments)
672,440 Million cell updates/sec

Title: US-09-412-100-23

Perfect score: 1 MSINTSGIAGSTWQISIGCA.....DAMAGDAINNALGKIGAA 403

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 197339 segs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCRTUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfilest1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2079	100.0	403	2	US-08-200-724A-2
2	2079	100.0	403	2	US-09-030-270A-3
3	2079	100.0	403	4	US-08-851-376A-2
4	2079	100.0	403	4	US-08-984-207-3
5	1928	92.7	385	1	US-08-891-254-3
6	1928	92.7	385	2	US-08-819-539-3
7	1928	92.7	385	5	PCT-US96-08819-3
8	1913	92.0	385	5	PCT-US93-06243-2
9	718.5	34.6	338	1	US-08-891-254-1
10	718.5	34.6	338	2	US-08-484-358-2
11	718.5	34.6	338	2	US-08-819-539-1
12	718.5	34.6	338	2	US-09-030-270A-1
13	718.5	34.6	338	3	US-09-118-959-2
14	718.5	34.6	338	4	US-08-984-207-1
15	718.5	34.6	338	5	PCT-US96-08819-1
16	211.5	10.2	318	4	US-09-060-756-727
17	211.5	10.2	718	1	US-08-425-069-2
18	211.5	10.2	718	2	US-08-317-844B-2
19	211.5	10.2	747	3	US-09-034-177-3
20	199.5	9.6	334	4	US-09-060-756-728
21	198.5	9.5	738	3	US-08-864-038A-3
22	197.5	9.5	344	1	US-08-891-254-7
23	197.5	9.5	344	2	US-08-819-539-7
24	197.5	9.5	344	2	US-09-030-270A-7
25	197.5	9.5	344	4	US-08-984-207-7
26	197.5	9.5	344	5	PCT-US96-08819-7
27	194	9.3	1160	3	US-08-808-599A-24

28	186	8.9	749	1	US-08-317-522A-2	Sequence 2, Appli
29	186	8.9	749	1	US-08-439-818A-2	Sequence 2, Appli
30	186	8.9	749	2	US-08-751-965-2	Sequence 2, Appli
31	186	8.9	749	2	US-08-738-975-2	Sequence 2, Appli
32	186	8.9	749	2	US-08-728-626-2	Sequence 2, Appli
33	186	8.9	749	3	US-08-808-599A-2	Sequence 2, Appli
34	185.5	8.8	235	2	US-08-529-108-1	Sequence 1, Appli
35	183	8.8	674	1	US-08-317-522A-3	Sequence 3, Appli
36	183	8.8	674	1	US-08-439-818A-3	Sequence 3, Appli
37	183	8.8	674	2	US-08-751-965-3	Sequence 3, Appli
38	183	8.8	674	2	US-08-738-975-3	Sequence 3, Appli
39	183	8.8	674	2	US-08-728-626-3	Sequence 3, Appli
40	183	8.8	674	3	US-08-808-599A-3	Sequence 3, Appli
41	171.5	8.2	745	2	US-09-010-928B-28	Sequence 28, Appli
42	171.5	8.2	870	2	US-09-010-928B-4	Sequence 2, Appli
43	171.5	8.2	907	2	US-09-010-928B-4	Sequence 4, Appli
44	171	8.2	1177	1	US-07-609-716-31	Sequence 31, Appli
45	171	8.2	1177	1	US-08-175-155-29	Sequence 29, Appli

ALIGNMENTS

RESULT 1
US-08-200-724A-2
Sequence 2, Application US/08200724A
Patent No. 5849868
GENERAL INFORMATION:
APPLICANT: Wei, Zhong-Min
APPLICANT: Bauer, David W.
APPLICANT: Beer, Steven V.
APPLICANT: Collier, Alan
APPLICANT: He, Sheng-Yang
APPLICANT: Laby, Ron J.
TITLE OF INVENTION: ELICITOR OF THE HYPERSENSITIVE RESPONSE
TITLE OF INVENTION: IN PLANTS
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon, Hargrave, Devans & Doyle
STREET: Clinton Square
CITY: Rochester
STATE: New York
COUNTRY: U.S.A.
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/200.724A
FILING DATE: 23-FEB-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Goldman, Michael L.
REGISTRATION NUMBER: 30,727
REFERENCE/DOCKET NUMBER: 19603/10030
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 263-1304
TELEFAX: (716) 263-1600
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 403 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-200-724A-2
Query Match 100.0% Score 2079; DB 2; Length 403;
Best Local Similarity 100.0%; Pred. No. 5.4e-175;
Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MSINTSGLASTMQISIGAGAGNNGLGSTRONAGLGSNLSALGIGGQNDVTYNOLAGLL 60
Db 1 MSINTSGLASTMQISIGAGAGNNGLGSTRONAGLGSNLSALGIGGQNDVTYNOLAGLL 60
QY 61 TGMAMMAMMAGGGGLMGCGGLGAGLGNGLGSGGGLGEBLSNALNDMLGSLNTLGSKGGNN 120
Db 61 TGMAMMAMMAGGGGLMGCGGLGAGLGNGLGSGGGLGEBLSNALNDMLGSLNTLGSKGGNN 120
QY 121 TTSTNSPLDQALGINSTONDDSTGTDSTSDSPMQLLMFSEIMOSLFGDGDGT 180
Db 121 TTSTNSPLDQALGINSTONDDSTGTDSTSDSPMQLLMFSEIMOSLFGDGDGT 180
QY 181 QGSSSGKOPTBEGBNAYKRGVTDALSGLMNGLSQLLGNGLGSGGAGNAGTGLDSSSL 240
Db 181 QGSSSGKOPTBEGBNAYKRGVTDALSGLMNGLSQLLGNGLGSGGAGNAGTGLDSSSL 240
QY 241 GKGGLQMLSGPVYQOLGNVGTGIGKAGIOALNDIGTRHSSSTRSFVNGDRAMAKEI 300
Db 241 GKGGLQMLSGPVYQOLGNVGTGIGKAGIOALNDIGTRHSSSTRSFVNGDRAMAKEI 300
QY 301 GOFMDQPEVFGKPYOKGPGOEKTDKSWAKALSKPDDGMPASMEQFNNAKGMIR 360
Db 301 GOFMDQPEVFGKPYOKGPGOEKTDKSWAKALSKPDDGMPASMEQFNNAKGMIR 360
QY 361 PMAGDTGNGMLQARGAGSSSLGIDAMMAGDAINNMALGKLGAA 403
Db 361 PMAGDTGNGMLQARGAGSSSLGIDAMMAGDAINNMALGKLGAA 403

RESULT 2
US-09-030-270A-3
; Sequence 3, Application us/09030270A
; Patent No. 5977060
; GENERAL INFORMATION:
; APPLICANT: Zitter, Thomas A.
; APPLICANT: Wei, Zhong-Min
; TITLE OF INVENTION: INSECT CONTROL WITH A
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Haigraue, Devans & Doyle LLP
; STREET: P.O. Box 1051, Clinton Square
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/030,270A
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/039,226
; FILING DATE: 28-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/1521
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEO ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 403 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein

```

```

US-09-030-270A-3
Query Match 100.0%; Score 2079; DB 2; Length 403;
Best Local Similarity 100.0%; Pred. No. 5,4e-175;
Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSINTSGLASTMQISIGAGAGNNGLGSTRONAGLGSNLSALGIGGQNDVTYNOLAGLL 60
Db 1 MSINTSGLASTMQISIGAGAGNNGLGSTRONAGLGSNLSALGIGGQNDVTYNOLAGLL 60
QY 61 TGMAMMAMMAGGGGLMGCGGLGAGLGNGLGSGGGLGEBLSNALNDMLGSLNTLGSKGGNN 120
Db 61 TGMAMMAMMAGGGGLMGCGGLGAGLGNGLGSGGGLGEBLSNALNDMLGSLNTLGSKGGNN 120
QY 121 TTSTNSPLDQALGINSTONDDSTGTDSTSDSPMQLLMFSEIMOSLFGDGDGT 180
Db 121 TTSTNSPLDQALGINSTONDDSTGTDSTSDSPMQLLMFSEIMOSLFGDGDGT 180
QY 181 QGSSSGKOPTBEGBNAYKRGVTDALSGLMNGLSQLLGNGLGSGGAGNAGTGLDSSSL 240
Db 181 QGSSSGKOPTBEGBNAYKRGVTDALSGLMNGLSQLLGNGLGSGGAGNAGTGLDSSSL 240
QY 241 GKGGLQMLSGPVYQOLGNVGTGIGKAGIOALNDIGTRHSSSTRSFVNGDRAMAKEI 300
Db 241 GKGGLQMLSGPVYQOLGNVGTGIGKAGIOALNDIGTRHSSSTRSFVNGDRAMAKEI 300
QY 301 GOFMDQPEVFGKPYOKGPGOEKTDKSWAKALSKPDDGMPASMEQFNNAKGMIR 360
Db 301 GOFMDQPEVFGKPYOKGPGOEKTDKSWAKALSKPDDGMPASMEQFNNAKGMIR 360
QY 361 PMAGDTGNGMLQARGAGSSSLGIDAMMAGDAINNMALGKLGAA 403
Db 361 PMAGDTGNGMLQARGAGSSSLGIDAMMAGDAINNMALGKLGAA 403

RESULT 3
US-08-851-376A-2
; Sequence 2, Application US/08851376A
; Patent No. 6174717
; GENERAL INFORMATION:
; APPLICANT: Beer, Steven V.
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Bauer, David W.
; APPLICANT: Collmer, Alan
; APPLICANT: He, Sheng-Yang
; APPLICANT: Lady, Ron
; TITLE OF INVENTION: ELICITOR OF THE HYPERSENSITIVE RESPONSE
; TITLE OF INVENTION: IN PLANTS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon Peabody LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/851,376A
; FILING DATE: 05-MAY-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/200,724
; FILING DATE: 23-FEB-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/10035

```

TELECOMMUNICATION INFORMATION:
 TELEPHONE: (716) 263-1304
 TELEFAX: (716) 263-1600
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 403 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 US-08-851-376A-2

Query Match 100.0%; Score 2079; DB 4; Length 403;
 Best Local Similarity 100.0%; Pred. No. 5.4e-175;
 Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLNTSGIGASTMQLISIGAGGNNGLTSTRONAGLGSNSALGLGGGNDTVNQLAGLL 60
 DB 1 MSLNTSGIGASTMQLISIGAGGNNGLTSTRONAGLGSNSALGLGGGNDTVNQLAGLL 60
 QY 61 TGMAMMSMGGGGLMGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGG 120
 DB 61 TGMAMMSMGGGGLMGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGG 120
 QY 121 TTSTNSPLDQALGINTSQNDSTSGTSTSDSSDPMQLKMFSEIMOSLFGDGQDGT 180
 DB 121 TTSTNSPLDQALGINTSQNDSTSGTSTSDSSDPMQLKMFSEIMOSLFGDGQDGT 180
 QY 181 QGSSSGGKOPTGEGONAKKVTYDALSGLMNGLSQLLNGGLGGGCGNAGTGLDGSSL 240
 DB 181 QGSSSGGKOPTGEGONAKKVTYDALSGLMNGLSQLLNGGLGGGCGNAGTGLDGSSL 240
 QY 241 GKGGLQNLSPDYQOLGNAVGTGIGMKAGIQALNDIGTHRSSSTRSVNKKDRAMAEI 300
 DB 241 GKGGLQNLSPDYQOLGNAVGTGIGMKAGIQALNDIGTHRSSSTRSVNKKDRAMAEI 300
 QY 301 GQFMQYPEVFEKPYQKPGQEVKTDKSWAKALSKPDDGMPASMEQFNKAKGMIR 360
 DB 301 GQFMQYPEVFEKPYQKPGQEVKTDKSWAKALSKPDDGMPASMEQFNKAKGMIR 360
 QY 361 PMAGDTGNLQARGAGSSSLGIDAMMAGDAINNALGKLGAA 403
 DB 361 PMAGDTGNLQARGAGSSSLGIDAMMAGDAINNALGKLGAA 403

RESULT 4

US-08-984-207-3
 Sequence 3, Application US/08984207
 Patent No. 6235974

GENERAL INFORMATION:
 APPLICANT: Qiu, Dewen
 APPLICANT: Wei, Zhong-Min
 APPLICANT: Beer, Steven V.
 TITLE OF INVENTION: HYPERSENSITIVE RESPONSE INDUCED
 TITLE OF INVENTION: RESISTANCE IN PLANTS BY SEED TREATMENT
 NUMBER OF SEQUENCES: 10
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
 STREET: P. O. Box 1051, Clinton Square
 CITY: Rochester
 STATE: New York
 COUNTRY: U.S.A.
 ZIP: 14603
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/984,207
 FILING DATE:
 CLASSIFICATION:

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 60/033,230
 FILING DATE: 05-DEC-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Goldman, Michael L.
 REGISTRATION NUMBER: 30,1727
 REFERENCE/DOCKET NUMBER: 19603/1201
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (716) 263-1304
 TELEFAX: (716) 263-1600
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 403 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-984-207-3

Query Match 100.0%; Score 2079; DB 4; Length 403;
 Best Local Similarity 100.0%; Pred. No. 5.4e-175;
 Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLNTSGIGASTMQLISIGAGGNNGLTSTRONAGLGSNSALGLGGGNDTVNQLAGLL 60
 DB 1 MSLNTSGIGASTMQLISIGAGGNNGLTSTRONAGLGSNSALGLGGGNDTVNQLAGLL 60
 QY 61 TGMAMMSMGGGGLMGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGG 120
 DB 61 TGMAMMSMGGGGLMGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGG 120
 QY 121 TTSTNSPLDQALGINTSQNDSTSGTSTSDSSDPMQLKMFSEIMOSLFGDGQDGT 180
 DB 121 TTSTNSPLDQALGINTSQNDSTSGTSTSDSSDPMQLKMFSEIMOSLFGDGQDGT 180
 QY 181 QGSSSGGKOPTGEGONAKKVTYDALSGLMNGLSQLLNGGLGGGCGNAGTGLDGSSL 240
 DB 181 QGSSSGGKOPTGEGONAKKVTYDALSGLMNGLSQLLNGGLGGGCGNAGTGLDGSSL 240
 QY 241 GKGGLQNLSPDYQOLGNAVGTGIGMKAGIQALNDIGTHRSSSTRSVNKKDRAMAEI 300
 DB 241 GKGGLQNLSPDYQOLGNAVGTGIGMKAGIQALNDIGTHRSSSTRSVNKKDRAMAEI 300
 QY 301 GQFMQYPEVFEKPYQKPGQEVKTDKSWAKALSKPDDGMPASMEQFNKAKGMIR 360
 DB 301 GQFMQYPEVFEKPYQKPGQEVKTDKSWAKALSKPDDGMPASMEQFNKAKGMIR 360
 QY 361 PMAGDTGNLQARGAGSSSLGIDAMMAGDAINNALGKLGAA 403
 DB 361 PMAGDTGNLQARGAGSSSLGIDAMMAGDAINNALGKLGAA 403

RESULT 5

US-08-891-254-3
 Sequence 3, Application US/08891254
 Patent No. 5776889

GENERAL INFORMATION:
 APPLICANT: Wei, Zhong-Min
 APPLICANT: Beer, Steven V.
 TITLE OF INVENTION: Hyperensitive Response
 TITLE OF INVENTION: Induced Resistance in Plants
 NUMBER OF SEQUENCES: 9
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Nixon, Hargrave, Devans & Doyle
 STREET: Clinton Square, P.O. Box 1051
 CITY: Rochester
 STATE: New York
 COUNTRY: U.S.A.
 ZIP: 14603
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/891,254
; FILING DATE: 10-JUL-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/475,775
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 14603/10050
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 385 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-891-254-3

```

```

Query Match          92.7%; Score 1928; DB 1; Length 385;
Best Local Similarity 100.0%; Pred. No. 9.4e-162;
Matches 372; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 MSNTSGLAGSTWQISIGAGAGNNGLLGTSRONAGLGNSALGAGGNQNDVTYNQLAGLL 60
DB 1 MSNTSGLAGSTWQISIGAGAGNNGLLGTSRONAGLGNSALGAGGNQNDVTYNQLAGLL 60
QY 61 TGMAMMMNMGGGGLMGGGGLGAGGGLGNGLGSGGGLGEGLSNALNDMLGGSLLNTLTSKGGNN 120
DB 61 TGMAMMMNMGGGGLMGGGGLGAGGGLGNGLGSGGGLGEGLSNALNDMLGGSLLNTLTSKGGNN 120
QY 121 TTSTNSPLDQALGINSQNDSTSGTSDTSDDPMQQLKMFSEIMOSLFGDGDGT 180
DB 121 TTSTNSPLDQALGINSQNDSTSGTSDTSDDPMQQLKMFSEIMOSLFGDGDGT 180
QY 121 TTSTNSPLDQALGINSQNDSTSGTSDTSDDPMQQLKMFSEIMOSLFGDGDGT 180
DB 121 TTSTNSPLDQALGINSQNDSTSGTSDTSDDPMQQLKMFSEIMOSLFGDGDGT 180
QY 181 QGSSSGGKQPTGEGONAYKKGVTDALSGLMNGLSQLLGNGLGAGGCGGAGNAGTGLDSSL 240
DB 181 QGSSSGGKQPTGEGONAYKKGVTDALSGLMNGLSQLLGNGLGAGGCGGAGNAGTGLDSSL 240
QY 241 GGGKGLQNLSCPVYDQOIGNAVGTGIGKAGIQALNDIGTRHHSSTRSFVNGDRAMAKEY 300
DB 241 GGGKGLQNLSCPVYDQOIGNAVGTGIGKAGIQALNDIGTRHHSSTRSFVNGDRAMAKEY 300
QY 301 GGFMDQYPEVFGKPOYQKPGQEVKTDKSWAKALSKRPDDGGMTPASMEOPNKAQMIKR 360
DB 301 GGFMDQYPEVFGKPOYQKPGQEVKTDKSWAKALSKRPDDGGMTPASMEOPNKAQMIKR 360
QY 361 PMAGDTGNGNLQ 372
DB 361 PMAGDTGNGNLQ 372

```

```

RESULT 6
US-08-819-539-3
; Sequence 3, Application US/08819539
; Patent No. 5859324
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Beer, Steven V.
; TITLE OF INVENTION: Hypersensitive Response
; TITLE OF INVENTION: Induced Resistance in Plants
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester

```

```

; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/819,539
; FILING DATE: 17-Mar-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/475,775
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 14603/10050
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 385 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-819-539-3

```

```

Query Match          92.7%; Score 1928; DB 2; Length 385;
Best Local Similarity 100.0%; Pred. No. 9.4e-162;
Matches 372; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 MSNTSGLAGSTWQISIGAGAGNNGLLGTSRONAGLGNSALGAGGNQNDVTYNQLAGLL 60
DB 1 MSNTSGLAGSTWQISIGAGAGNNGLLGTSRONAGLGNSALGAGGNQNDVTYNQLAGLL 60
QY 61 TGMAMMMNMGGGGLMGGGGLGAGGGLGNGLGSGGGLGEGLSNALNDMLGGSLLNTLTSKGGNN 120
DB 61 TGMAMMMNMGGGGLMGGGGLGAGGGLGNGLGSGGGLGEGLSNALNDMLGGSLLNTLTSKGGNN 120
QY 121 TTSTNSPLDQALGINSQNDSTSGTSDTSDDPMQQLKMFSEIMOSLFGDGDGT 180
DB 121 TTSTNSPLDQALGINSQNDSTSGTSDTSDDPMQQLKMFSEIMOSLFGDGDGT 180
QY 121 TTSTNSPLDQALGINSQNDSTSGTSDTSDDPMQQLKMFSEIMOSLFGDGDGT 180
DB 121 TTSTNSPLDQALGINSQNDSTSGTSDTSDDPMQQLKMFSEIMOSLFGDGDGT 180
QY 181 QGSSSGGKQPTGEGONAYKKGVTDALSGLMNGLSQLLGNGLGAGGCGGAGNAGTGLDSSL 240
DB 181 QGSSSGGKQPTGEGONAYKKGVTDALSGLMNGLSQLLGNGLGAGGCGGAGNAGTGLDSSL 240
QY 241 GGGKGLQNLSCPVYDQOIGNAVGTGIGKAGIQALNDIGTRHHSSTRSFVNGDRAMAKEY 300
DB 241 GGGKGLQNLSCPVYDQOIGNAVGTGIGKAGIQALNDIGTRHHSSTRSFVNGDRAMAKEY 300
QY 301 GGFMDQYPEVFGKPOYQKPGQEVKTDKSWAKALSKRPDDGGMTPASMEOPNKAQMIKR 360
DB 301 GGFMDQYPEVFGKPOYQKPGQEVKTDKSWAKALSKRPDDGGMTPASMEOPNKAQMIKR 360
QY 361 PMAGDTGNGNLQ 372
DB 361 PMAGDTGNGNLQ 372

```

```

RESULT 7
PCT-US96-08819-3
; Sequence 3, Application PC/TUS9608819
; GENERAL INFORMATION:
; APPLICANT: Cornell Research Foundation, Inc.
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE INDUCED
; TITLE OF INVENTION: RESISTANCE IN PLANTS
; NUMBER OF SEQUENCES: 9

```

CORRESPONDENCE ADDRESS:
 ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
 STREET: Clinton Square, P.O. Box 1051
 CITY: Rochester
 STATE: New York
 COUNTRY: U.S.A.
 ZIP: 14603
 COMPUTER READABLE FORM:
 MEDIUM TYPE: floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US96/08819
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/475,775
 FILING DATE: 07-JUN-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Goldman, Michael L.
 REGISTRATION NUMBER: 30,727
 REFERENCE/DOCKET NUMBER: 19603/10051
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (716) 263-1304
 TELEFAX: (716) 263-1600
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 385 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 PCT-US96-08819-3

Query Match 92.7%; Score 1928; DB 5; Length 385;
 Best Local Similarity 100.0%; Pred. No. 9,4e-162;
 Matches 372; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLNSTSGLAGASTMOJISIGAGGNNGLGTSRONAGIGNSALGIGGNNNDTVNOLAGLL 60
 DB 1 MSLNSTSGLAGASTMOJISIGAGGNNGLGTSRONAGIGNSALGIGGNNNDTVNOLAGLL 60
 QY 61 TGMNMMMSMGGGGLMGGLGGGLGNGLGSGGLGEGLSNALNDMLGSLNTLGSKGGNN 120
 DB 61 TGMNMMMSMGGGGLMGGLGGGLGNGLGSGGLGEGLSNALNDMLGSLNTLGSKGGNN 120
 QY 121 TTSTTNSPLDQALGINSTSQNDSTSGTDTSDSSDPMOQLKMFSEIMQSLFGGQDGT 180
 DB 121 TTSTTNSPLDQALGINSTSQNDSTSGTDTSDSSDPMOQLKMFSEIMQSLFGGQDGT 180
 QY 181 QSSSSGGKOPTGEBONAYKKYTDALSGLMNGLSQLLNGGLGGGCGGNAAGTGLDSSSL 240
 DB 181 QSSSSGGKOPTGEBONAYKKYTDALSGLMNGLSQLLNGGLGGGCGGNAAGTGLDSSSL 240
 QY 241 GKGGLQNLNSGPVYDQOLGNAVGTGIGMKAGIOALNDIGTHRSSSTRSFVNKGDRAKAKET 300
 DB 241 GKGGLQNLNSGPVYDQOLGNAVGTGIGMKAGIOALNDIGTHRSSSTRSFVNKGDRAKAKET 300
 QY 301 GQPMDOYREVEFGKPOYOKPGQGEVKTDDKSWAKALSKPDDGMPASMOQFNKAKGMIKR 360
 DB 301 GQPMDOYREVEFGKPOYOKPGQGEVKTDDKSWAKALSKPDDGMPASMOQFNKAKGMIKR 360
 QY 361 PMAGDTGNGNL 372
 DB 361 PMAGDTGNGNL 372

RESULT 8
 PCT-US93-06243--2
 Sequence 2, Application PC/TUS9306243
 GENERAL INFORMATION:

APPLICANT: Zhong-Min Wei, David M. Bauer, Steven V.
 APPLICANT: Beer, Alan Collmer, Sheng-Yang He, and Ron J. Laby
 TITLE OF INVENTION: Elicitor of the Hypersensitive Response in Plants
 NUMBER OF SEQUENCES: 5
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Yahwak & Associates
 STREET: 25 Skytop Drive
 CITY: Trumbull
 STATE: Connecticut
 COUNTRY: USA
 ZIP: 06611
 COMPUTER READABLE FORM:
 MEDIUM TYPE: floppy disk
 COMPUTER: Macintosh
 OPERATING SYSTEM: MS-DOS
 SOFTWARE: Microsoft Word 4.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US93/06243
 FILING DATE: 19930630
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 907,935
 FILING DATE: 01-JUL-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: George M. Yahwak
 REGISTRATION NUMBER: 26,824
 REFERENCE/DOCKET NUMBER: CRF D-1172
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (203)268-1951
 TELEFAX: (203)268-1951
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 385 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 PCT-US93-06243-2

Query Match 92.0%; Score 1913; DB 5; Length 385;
 Best Local Similarity 99.5%; Pred. No. 1.9e-160;
 Matches 369; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSLNSTSGLAGASTMOJISIGAGGNNGLGTSRONAGIGNSALGIGGNNNDTVNOLAGLL 60
 DB 1 MSLNSTSGLAGASTMOJISIGAGGNNGLGTSRONAGIGNSALGIGGNNNDTVNOLAGLL 60
 QY 61 TGMNMMMSMGGGGLMGGLGGGLGNGLGSGGLGEGLSNALNDMLGSLNTLGSKGGNN 120
 DB 61 TGMNMMMSMGGGGLMGGLGGGLGNGLGSGGLGEGLSNALNDMLGSLNTLGSKGGNN 120
 QY 121 TTSTTNSPLDQALGINSTSQNDSTSGTDTSDSSDPMOQLKMFSEIMQSLFGGQDGT 180
 DB 121 TTSTTNSPLDQALGINSTSQNDSTSGTDTSDSSDPMOQLKMFSEIMQSLFGGQDGT 180
 QY 181 QSSSSGGKOPTGEBONAYKKYTDALSGLMNGLSQLLNGGLGGGCGGNAAGTGLDSSSL 240
 DB 181 QSSSSGGKOPTGEBONAYKKYTDALSGLMNGLSQLLNGGLGGGCGGNAAGTGLDSSSL 240
 QY 241 GKGGLQNLNSGPVYDQOLGNAVGTGIGMKAGIOALNDIGTHRSSSTRSFVNKGDRAKAKET 300
 DB 241 GKGGLQNLNSGPVYDQOLGNAVGTGIGMKAGIOALNDIGTHRSSSTRSFVNKGDRAKAKET 300
 QY 301 GQPMDOYREVEFGKPOYOKPGQGEVKTDDKSWAKALSKPDDGMPASMOQFNKAKGMIKR 360
 DB 301 GQPMDOYREVEFGKPOYOKPGQGEVKTDDKSWAKALSKPDDGMPASMOQFNKAKGMIKR 360
 QY 361 PMAGDTGNGNL 371
 DB 361 PMAGDTGNGNL 371

QY 300 IGOFPMDQPEYFEGKPYOKGQGEVKTDDKSMARALSKPDDGKTPASMEOPNKAQK 359
Db 235 IGOFPMDQPEYFEGKPYOKGQGEVKTDDKSMARALSKPDDGKTPASMEOPNKAQK 294
QY 360 RPMAGDGTGNTNLNRGAGASLGIDAAVYGDKIANNMISGLKLANA 403
Db 295 SAVAGDGTGNTNLNRGAGASLGIDAAVYGDKIANNMISGLKLANA 338

RESULT 11

US-08-819-539-1
Sequence 1, Application US/08819539
Patent No. 5859324
GENERAL INFORMATION:
APPLICANT: Wei, Zhong-Min
APPLICANT: Beer, Steven V.
TITLE OF INVENTION: Hypersensitive Response
TITLE OF INVENTION: Induced Resistance in Plants
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon, Hargrave, Devans & Doyle
STREET: Clinton Square, P.O. Box 1051
CITY: Rochester
STATE: New York
COUNTRY: U.S.A.
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/819,539
FILING DATE: 17-MAR-1997
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/475,775
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Goldman, Michael L.
REGISTRATION NUMBER: 30,727
REFERENCE/DOCKET NUMBER: 14603/10050
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 263-1304
TELEFAX: (716) 263-1600
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 338 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-819-539-1

Query Match 34.6%; Score 718.5; DB 2; Length 338;
Best Local Similarity 42.8%; Pred. No. 1.4e-55;
Matches 173; Conservative 41; Mismatches 111; Indels 79; Gaps 11;

QY 13 MOIST-GGAGGNNGLGTSRONAGIGG-NSA---LGLGGGNDNTVNOGLAGLLTGMNMM 67
Db 1 MOITIKAHIGDLGVSGAGQ--GLKGLNSAASSIGSSVDKLSITDKLTSALTSNM--- 55
QY 68 SMAGGGGLMGGGLGGLGNGLGSGGLGEGLSNALNDMLGSLNLTGSKGCGNNTSTINS 127
Db 56 -----FGGALAOGIGLAS--SKGIGMSNOLGOSFGN-----GAQAGASNLSVPR- 96
QY 128 PLDQALGINSTQNDSTSGTSDSDPMDQLKMFSEIMQSLFG-----DGQDG 179
Db 97 -----SGGDALS-----KMFDAKADLDLGHDTYKRLNQSNQ 128
QY 180 TGGSSGGKQPTGEGONAYKKGVTDALSLGMNGLSQLLGNGLGGGCGGNGAGTGLDSS 239

Db 129 LANSMLNASOMTQGNMNAFSGGVNNAALSYINGNIGQSM-----SGFSOPS 174
QY 240 LGKGLGILNSGPVVOYQGNVGTGIGKACIGQALNDIGTRHSSTSFYVKKGRAMAKE 299
Db 175 LGAGIGLGSAGAFNOLGNAIGMGVGNALSLNSVSTHVDGNNNRHFVDKEDRGNAKE 234
QY 300 IGOFPMDQPEYFEGKPYOKGQGEVKTDDKSMARALSKPDDGKTPASMEOPNKAQK 359
Db 235 IGOFPMDQPEYFEGKPYOKGQGEVKTDDKSMARALSKPDDGKTPASMEOPNKAQK 294
QY 360 RPMAGDGTGNTNLNRGAGASLGIDAAVYGDKIANNMISGLKLANA 403
Db 295 SAVAGDGTGNTNLNRGAGASLGIDAAVYGDKIANNMISGLKLANA 338

RESULT 12

US-09-030-270A-1
Sequence 1, Application US/09030270A
Patent No. 597060
GENERAL INFORMATION:
APPLICANT: Zitter, Thomas A.
APPLICANT: Wei, Zhong-Min
TITLE OF INVENTION: INSECT CONTROL WITH A
TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
STREET: P.O. Box 1051, Clinton Square
CITY: Rochester
STATE: New York
COUNTRY: U.S.A.
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/030,270A
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/039,226
FILING DATE: 28-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Goldman, Michael L.
REGISTRATION NUMBER: 30,727
REFERENCE/DOCKET NUMBER: 19603/1521
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 263-1304
TELEFAX: (716) 263-1600
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 338 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-030-270A-1

Query Match 34.6%; Score 718.5; DB 2; Length 338;
Best Local Similarity 42.8%; Pred. No. 1.4e-55;
Matches 173; Conservative 41; Mismatches 111; Indels 79; Gaps 11;

QY 13 MOIST-GGAGGNNGLGTSRONAGIGG-NSA---LGLGGGNDNTVNOGLAGLLTGMNMM 67
Db 1 MOITIKAHIGDLGVSGAGQ--GLKGLNSAASSIGSSVDKLSITDKLTSALTSNM--- 55
QY 68 SMAGGGGLMGGGLGGLGNGLGSGGLGEGLSNALNDMLGSLNLTGSKGCGNNTSTINS 127
Db 56 -----FGGALAOGIGLAS--SKGIGMSNOLGOSFGN-----GAQAGASNLSVPR- 96

QY 128 PLDQALGINSTQNDSTGSDTSDSSDPMOQLKMFSEIMOGLFG-----DGODG 179
DB 97 -----SGGDALS-----KMFDRKALDDLLGHDVYTKLTLNQSNO 128
QY 180 TQSSSGGKOPTBEONAYKKGYVDALSGLMGNLSOLLGNGGAGGOGGAGTGLDSS 239
DB 129 LANSMLNASQNTQGNMNAFSGGVNNAALSLILGNGLGQSM-----SGFSQPS 174
QY 240 LGGKGLQNLGSPVDYQOOLGNVAVGTGIGKAGIQALNDIGTHRSSTSFVYKGRANAKE 299
DB 175 LGAGGLQGLSGAGAFNQLGNALIGMGVGNALSLSNVSTHVDCNNHFFYDKEDRGMAKE 234
QY 300 IGFPMDOYPEVFGKPYOKGPGQEVKTDKSKWAKALSKPDDGWTSPASMEQFNKAKGMK 359
DB 235 IGFPMDOYPEIFGKPEYOKDGWSSPKTDDKSKWAKALSKPDDGWTGASMDKFRQAMGMK 294
QY 360 RPMAGDTGNGNLQARGAGSSIGIDAMAGDAINNMALGKIGAA 403
DB 295 SAVAGDTGNTNLNRGAGASLGIDAAYGDKIANMSLIGKILANA 338
RESULT 13
US-09-118-959-2
; Sequence 2, Application US/09118959
; Patent No. 6001959
; GENERAL INFORMATION:
; APPLICANT: Bauer, David
; APPLICANT: Collmer, Alan
; TITLE OF INVENTION: Hypersensitive Response Elicitor From
; TITLE OF INVENTION: Erythrina Chrysanthem
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/118,959
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldmen, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/840
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 716-263-1304
; TELEFAX: 716-263-1600
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 338 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-118-959-2
Query Match 34.6%; Score 718.5; DB 3; Length 338;
Best Local Similarity 42.8%; Pred. No. 1,4e-55;
Matches 1/3; Conservative 41; Mismatches 111; Indels 79; Gaps 11;
QY 13 MOISI-GGAGGNGGLGTSRONMGILG-NSA---IGLGGGQNTVQNLGALLTGNNMM 67
DB 1 MOTTIRAHIGDLGVSLGAG---GLKGLNSAASLGSSVDKLSITDKTSALTSMM---55

QY 68 SMGCGGLMGGLGGLGNGLGSSGLGEGLSNALNDMLTGSGLNTLTKSGKGNNTTSTTNS 127
DB 56 -----FGALAGIGLIGAS-SKGLGMSNOLGGSFGN-----GAQASNLISLPK- 96
QY 128 PLDQALGINSTQNDSTGSDTSDSSDPMOQLKMFSEIMOGLFG-----DGODG 179
DB 97 -----SGGDALS-----KMFDRKALDDLLGHDVYTKLTLNQSNO 128
QY 180 TQSSSGGKOPTBEONAYKKGYVDALSGLMGNLSOLLGNGGAGGOGGAGTGLDSS 239
DB 129 LANSMLNASQNTQGNMNAFSGGVNNAALSLILGNGLGQSM-----SGFSQPS 174
QY 240 LGGKGLQNLGSPVDYQOOLGNVAVGTGIGKAGIQALNDIGTHRSSTSFVYKGRANAKE 299
DB 175 LGAGGLQGLSGAGAFNQLGNALIGMGVGNALSLSNVSTHVDCNNHFFYDKEDRGMAKE 234
QY 300 IGFPMDOYPEVFGKPYOKGPGQEVKTDKSKWAKALSKPDDGWTSPASMEQFNKAKGMK 359
DB 235 IGFPMDOYPEIFGKPEYOKDGWSSPKTDDKSKWAKALSKPDDGWTGASMDKFRQAMGMK 294
QY 360 RPMAGDTGNGNLQARGAGSSIGIDAMAGDAINNMALGKIGAA 403
DB 295 SAVAGDTGNTNLNRGAGASLGIDAAYGDKIANMSLIGKILANA 338
RESULT 14
US-08-984-207-1
; Sequence 1, Application US/08984207
; Patent No. 623574
; GENERAL INFORMATION:
; APPLICANT: Qiu, Dewen
; APPLICANT: Wei, Zhong-Min
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE INDUCED
; TITLE OF INVENTION: RESISTANCE IN PLANTS BY SEED TREATMENT
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: P.O. Box 1051, Clinton Square
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/984,207
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/033,230
; FILING DATE: 05-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/1201
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 338 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-984-207-1
Query Match 34.6%; Score 718.5; DB 4; Length 338;

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 25, 2001, 11:43:24 ; Search time 22.79 Seconds
(without alignments)
1072.026 Million cell updates/sec

Title: US-09-412-100-23

Perfect score: 2079
Sequence: 1 MSINISGLASTWQISIGGA.....DAMAGDAINNALGLGAA 403

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

1: /SID58/gcgdata/geneseq/AA1980.DAT:*
2: /SID58/gcgdata/geneseq/AA1981.DAT:*
3: /SID58/gcgdata/geneseq/AA1982.DAT:*
4: /SID58/gcgdata/geneseq/AA1983.DAT:*
5: /SID58/gcgdata/geneseq/AA1984.DAT:*
6: /SID58/gcgdata/geneseq/AA1985.DAT:*
7: /SID58/gcgdata/geneseq/AA1986.DAT:*
8: /SID58/gcgdata/geneseq/AA1987.DAT:*
9: /SID58/gcgdata/geneseq/AA1988.DAT:*
10: /SID58/gcgdata/geneseq/AA1989.DAT:*
11: /SID58/gcgdata/geneseq/AA1990.DAT:*
12: /SID58/gcgdata/geneseq/AA1991.DAT:*
13: /SID58/gcgdata/geneseq/AA1992.DAT:*
14: /SID58/gcgdata/geneseq/AA1993.DAT:*
15: /SID58/gcgdata/geneseq/AA1994.DAT:*
16: /SID58/gcgdata/geneseq/AA1995.DAT:*
17: /SID58/gcgdata/geneseq/AA1996.DAT:*
18: /SID58/gcgdata/geneseq/AA1997.DAT:*
19: /SID58/gcgdata/geneseq/AA1998.DAT:*
20: /SID58/gcgdata/geneseq/AA1999.DAT:*
21: /SID58/gcgdata/geneseq/AA2000.DAT:*
22: /SID58/gcgdata/geneseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2079	100.0	403	AAW75863	Erwinia amylovora
2	2079	100.0	403	AAW62455	Erwinia amylovora
3	2079	100.0	403	AAW61114	Hypersensitive res
4	2079	100.0	403	AAW87639	A hypersensitive r
5	2079	100.0	403	AAV71093	Erwinia amylovora
6	2079	100.0	403	AAV84854	A hypersensitive r
7	1928	92.7	385	AAW06598	Hypersensitive res
8	1913	92.0	385	AAW45751	Erwinia amylovora
9	718.5	34.6	338	AAW05597	Hypersensitive res
10	718.5	34.6	338	AAW87638	A hypersensitive r
11	718.5	34.6	338	AAW82407	E. chrysanthemi hr

12	718.5	34.6	338	21	AAV71092
13	718.5	34.6	338	21	AAV84853
14	718.5	34.6	338	21	AAV55801
15	718.5	34.6	340	19	AAW75862
16	718.5	34.6	340	19	AAW61113
17	713.5	34.3	340	19	AAW62454
18	226.5	10.9	898	18	AAW31853
19	215.5	10.4	718	12	AAW41308
20	211.5	10.2	651	20	AAV40097
21	211.5	10.2	718	19	AAW53346
22	211.5	10.2	718	21	AAV59070
23	206	9.9	604	16	AAW99057
24	200	9.6	388	20	AAV04999
25	199.5	9.6	731	21	AAW31750
26	198.5	9.5	708	19	AAW56163
27	198	9.5	646	18	AAW21178
28	197.5	9.5	344	18	AAW06600
29	197.5	9.5	344	19	AAW75865
30	197.5	9.5	344	19	AAW62457
31	197.5	9.5	344	19	AAW61116
32	197.5	9.5	344	20	AAW87641
33	197.5	9.5	344	21	AAV71099
34	197.5	9.5	344	21	AAW84860
35	197	9.5	606	16	AAW99055
36	197	9.5	606	20	AAV40101
37	197	9.5	606	20	AAV40102
38	194	9.3	1160	21	AAW23367
39	193.5	9.3	174	21	AAW39747
40	192.5	9.3	606	16	AAW99053
41	192.5	9.3	606	20	AAV40100
42	190	9.1	641	20	AAV28843
43	190	9.1	641	21	AAV95856
44	186	8.9	749	17	AAW94895
45	186	8.9	749	21	AAW23357

ALIGNMENTS

RESULT 1

AAW75863	Standard; Protein; 403 AA.
ID	AAW75863
AC	AAW75863;
XX	
DT	07-DEC-1998 (first entry)
XX	
DE	Erwinia amylovora hypersensitive response elicitor (HRE).
XX	
KW	Hypersensitive response elicitor; HRE; insect resistance;
KW	biological control; transgenic plant.
XX	
OS	Erwinia amylovora.
XX	
PN	W09837752-A1.
XX	
PD	03-SEP-1998.
XX	
PF	26-FEB-1998; 98MO-US03604.
XX	
PR	28-FEB-1997; 97US-0039226.
XX	
PA	(CORR) CORNELL RES FOUND INC.
XX	
PI	Wei Z, Zitter TA;
XX	
DR	WPI; 1998-495374/42.
XX	
DR	N-PSDB; AAV54607.
XX	
PT	use of hypersensitive response elicitor polypeptide - for
PT	application to plants or seeds or transgenic plants or seeds for the
XX	control of insects.

Erwinia chrysanthemi
A hypersensitive r
E. chrysanthemi hy
Erwinia chrysanth
Hypersensitive res
Erwinia chrysanth
Mycobacterium tube
N. clavipes draglin
Spider silk protei
Nephila clavipes s
N. clavipes spider
Spider dragline va
Mycobacterium spec
Arabidopsis thalia
New DNA sequence 1
Nephila clavipes s
Hypersensitive res
Pseudomonas solana
Pseudomonas solana
Hypersensitive res
A hypersensitive r
Pseudomonas solana
A hypersensitive r
Spider dragline va
Polymer of an anal
Polymer of an anal
Polymer of an anal
Mouse tropinlin pr
Arabidopsis thalia
Spider dragline va
Polymer of an anal
Epstein Barr Virus
Epstein Barr Virus
Human tropinlin.
Human tropinlin pr

PS Disclosure; Page 9-10; 75pp; English.

XX This is the amino acid sequence of a 39 kDa, heat stable
 CC hypersensitive response elicitor (HRE) of *Erwinia amylovora*. The
 CC invention relates to the use of a HRE polypeptide or protein to
 CC control insects on plants or plants grown from seed treated with HRE.
 CC Also claimed is a method of insect control for plants that involves:
 CC (a) providing a transgenic plant or seed transformed with a DNA
 CC molecule (see AAV54606-09) encoding a HRE polypeptide or protein (see
 CC AAV75862-67); and (b) growing the transgenic plants or transgenic
 CC plants produced from the transgenic seeds to control insects. HRE
 CC prevents direct insect damage to plants by feeding injury. It kills
 CC insects close to plants, and interferes with insect larval feeding
 CC on such plants. It also prevents insects from colonising host
 CC plants and releasing phytotoxins which result in disease damage to
 CC plants.

SQ Sequence 403 AA;

Query Match 100.0%; Score 2079; DB 19; Length 403;
 Best Local Similarity 100.0%; Pred. No. 9.6e-156;
 Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSINTSGLAGSTWQISIGAGAGNNGILGTSRONAGLGGNSALGLGGGNDYVNOIAGLL 60
 DB 1 msintsglagstwmqisigagagnngilgtsrqnaglggnsalglgggnqndvtnqiaql 60
 QY 61 TGGMMAMNMMWGGGGLMGCGGLGNGLGSSGGLGEGLSNALMDMGGSLNTGSKGNN 120
 DB 61 tggmmamnmwgggglmgcgglgnglgssgglegelsnaldmaggslntgskgynn 120
 QY 121 TTSTNSPLDOALGINSTSONDSTSGTSDSPDMOOLKMFSEIMOSLFGDGODGT 180
 DB 121 ttstnspldoalginstsondstsgtstdspdmoolkmfseimoslfgdgodgt 180
 QY 122 tttstnspldqalginstsondstsgtstdspdmqllkmfseimgslfgdgodgt 180
 DB 122 tttstnspldqalginstsondstsgtstdspdmqllkmfseimgslfgdgodgt 180
 QY 181 QGSSSGGKOPTBEGBNAYKKGVTDAISGLMNGLSQLLNGSLGGGCGAGTGLDSS 240
 DB 181 qgsssggkoptbeganaykkgvtcdalsglnqnglsqllngsglgggagagtcgldgssl 240
 QY 241 GGGKGLNLSGPVYQOLGNVGTGIGKAGIOALNDIGTHRSSTRSFVKKGRRAAKET 300
 DB 241 gggkglnlsgpvdyqolgnvgtgigkagioalndigthrsstrsfvkkgramaekel 300
 QY 301 GGFMDQYPEVFGKPYOKGPGOEKTDKSMARALSKPDDGWTSPASMEQFNKAKGMIKR 360
 DB 301 ggfmdqypcvfgykpyqkpggdevktcdkswakalskppddgwtspasmeqfnkagmikt 360
 QY 361 PMAGDTGNGMLQARGAGSSSLGIDAMMAGDAINNMALGKLGAA 403
 DB 361 pmagdtgngnlqargagssslgidammagdainnmalgkligaa 403

RESULT 2

AAW62455 standard; Protein; 403 AA.

AAW62455;

09-NOV-1998 (first entry)

Erwinia amylovora hypersensitive response elicitor (HRE).

Hypersensitive response elicitor; HRE; growth; transgenic plant.

Erwinia amylovora.

WO9832844-A1.

30-JUL-1998.

27-JAN-1998; 98MO-US01507.

PR 27-JAN-1997; 97US-0036048.

XX (CORR) CORNELL RES FOUND INC.

PA Beer SV, Qiu D, Wei Z;

PI Beer SV, Qiu D, Wei Z;

XX WPI: 1998-427940/36.

DR N-PSDB; AAV39973.

XX Method for enhancing plant growth - comprises use of hypersensitive
 PT response elicitor polypeptide or protein which may also effect, e.g.
 PT increase in plant height or earlier germination seed

PS Disclosure; Page 15-16; 110pp; English.

CC This is the deduced amino acid sequence of the 39 kDa, heat stable
 CC hypersensitive response elicitor (HRE) of *Erwinia amylovora*. A
 CC method of enhancing growth in plants comprises: (a) applying a HRE
 CC polypeptide or protein in a non-infectious form to a plant or plant
 CC seed under conditions effective to enhance growth of the plant or
 CC plants grown from the seed, or (b) providing a transgenic plant or
 CC plant seed transformed with a DNA molecule encoding a HRE
 CC polypeptide or protein, and growing the transgenic plant or a plant
 CC produced from the transgenic seed under conditions effective to
 CC enhance plant growth. HRES (see AAW62454-59) or nucleic acids
 CC encoding them (see AAV39972-75) can be used to increase plant growth.
 CC The HRES may also result in increased plant height and yield, and
 CC effect early germination and maturation of plant seed and early
 CC coloration of fruit and plants. E. amylovora HRE can be applied
 CC to tomato plants to enhance growth without causing disease in that
 CC species; this bacterium is a pathogen of apple and pear but not
 CC of tomato.

SQ Sequence 403 AA;

Query Match 100.0%; Score 2079; DB 19; Length 403;
 Best Local Similarity 100.0%; Pred. No. 9.6e-156;
 Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSINTSGLAGSTWQISIGAGAGNNGILGTSRONAGLGGNSALGLGGGNDYVNOIAGLL 60
 DB 1 msintsglagstwmqisigagagnngilgtsrqnaglggnsalglgggnqndvtnqiaql 60
 QY 61 TGGMMAMNMMWGGGGLMGCGGLGNGLGSSGGLGEGLSNALMDMGGSLNTGSKGNN 120
 DB 61 tggmmamnmwgggglmgcgglgnglgssgglegelsnaldmaggslntgskgynn 120
 QY 121 TTSTNSPLDOALGINSTSONDSTSGTSDSPDMOOLKMFSEIMOSLFGDGODGT 180
 DB 121 tttstnspldoalginstsondstsgtstdspdmoolkmfseimoslfgdgodgt 180
 QY 122 tttstnspldqalginstsondstsgtstdspdmqllkmfseimgslfgdgodgt 180
 DB 122 tttstnspldqalginstsondstsgtstdspdmqllkmfseimgslfgdgodgt 180
 QY 181 QGSSSGGKOPTBEGBNAYKKGVTDAISGLMNGLSQLLNGSLGGGCGAGTGLDSS 240
 DB 181 qgsssggkoptbeganaykkgvtcdalsglnqnglsqllngsglgggagagtcgldgssl 240
 QY 241 GGGKGLNLSGPVYQOLGNVGTGIGKAGIOALNDIGTHRSSTRSFVKKGRRAAKET 300
 DB 241 gggkglnlsgpvdyqolgnvgtgigkagioalndigthrsstrsfvkkgramaekel 300
 QY 301 GGFMDQYPEVFGKPYOKGPGOEKTDKSMARALSKPDDGWTSPASMEQFNKAKGMIKR 360
 DB 301 ggfmdqypcvfgykpyqkpggdevktcdkswakalskppddgwtspasmeqfnkagmikt 360
 QY 361 PMAGDTGNGMLQARGAGSSSLGIDAMMAGDAINNMALGKLGAA 403
 DB 361 pmagdtgngnlqargagssslgidammagdainnmalgkligaa 403

RESULT 3

AAW61114 standard; Protein; 403 AA.

AAW61114

AA61114;
 26-OCT-1998 (first entry)
 Hypersensitive response elicitor protein (39 kDa).
 Hypersensitive response elicitor; transgenic plant; seed;
 pathogen resistance; disease resistance; crop protection.
 Erwinia amylovora.
 W09824297-A1.
 11-JUN-1998.
 04-DEC-1997; 97WO-US22629.
 05-DEC-1996; 96US-0033230.
 (CORR) CORNELL RES FOUND INC.
 Beer SV, Qiu D, Wei Z;
 WPI: 1998-332931/29.
 N-PSDB; AAV36428.
 Imparting pathogen resistance to plants. - by applying a
 hypersensitive response elicitor polypeptide to seeds
 Disclosure; Page 18-20; 85pp; English.
 This is the 39 kDa hypersensitive elicitor (HRE) protein of
 Erwinia amylovora. It is heat stable at 100 degC for at least 10
 min, has a pI of approximately 4.3, and contains substantially no
 cysteine. The invention relates to methods of imparting
 hypersensitive response induced resistance to plants by treatment
 of seeds. Isolated HRE proteins can be applied to seeds as a means
 of imparting pathogen resistance to plants grown from the seeds.
 Alternatively, bacteria containing the gene encoding the HRE can be
 applied to the plant seeds, or transgenic plant seeds containing a
 DNA molecule encoding an HRE polypeptide or protein are used. HRE
 polypeptide sequences from Erwinia chrysanthemi, Erwinia amylovora,
 Pseudomonas syringae, Pseudomonas solanacearum, Xanthomonas
 campestris pv. glycines and Xanthomonas campestris pelargonii (see
 AA61113-18) are provided. The methods can impart pathogen
 resistance without using agents which are harmful to the
 environment or pathogenic to the plant seed being treated, or to
 adjacent plants. E. amylovora causes disease in apple or pear
 but not tomato. However, it elicits a hypersensitive response in
 tomato. Thus, E. amylovora can be applied to tomato seeds to
 impart pathogen resistance without causing diseases in plants of
 that species.
 Sequence 403 AA:
 Query Match 100.0%; Score 2079; DB 19; Length 403;
 Best Local Similarity 100.0%; Pred. No. 9.6e-156;
 Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 1 MSLNTSGLAGASTWQISIGAGGNNGLGTSRONAGLGNSALGLGGNONDTVNLAGLL 60
 Db 1 mslntsglagastwqisigaggnngllgtsrqnaglgnsalglggnqndvtnqagll 60
 Qy 61 TGMNMMMSMGGGGGLAGGGLGNGLAGSGGLGBGLSNALNDMLGGLNTLGSKGNN 120
 Db 61 tgmnmmsmggggglagggllgnglagsggglgbglsnalnndmlgglnltlsgkgnn 120
 Qy 121 TTSTNSPLDQALGINTSTSONDSTSGTSDTSSDPMOQLKMFSEIMQSLFGGKODCT 180
 Db 121 ttstnspldqalginstsondstsgtstsdssdpmqllkmfseimqslfgggqdgct 180
 Qy 181 OGSSSGKQPTBGEDONAYKKVTDALSGLGNGLSQLLGLNGLGGGGGNAGTGLDGSSTL 240

181 gssssgkqptegeqnaaykkytdalsglnqslglnqslg9999naagctldgssl 240
 Qy 241 GKGKLNLSGPDVQOQLGNAVGTCIGMKAGTIOALNDICTGHHSSSTRSPVNGDRMAKEI 300
 Db 241 gkgklnlsdpdyqqlgnaavgctigmkagtdioalndictghhssstrspvngdrmakei 300
 Qy 301 GOFMDQYPEVEFGKPOYOKPGQEVKTDKSWAKALSKPDDQMPASMEOPNKAKGMTR 360
 Db 301 gfmddqypevfkgpyqgpggevtddkswakalskpdddgmprasmegfnkakgmtr 360
 Qy 361 PMAGDTGKNTQARCGAGSSSLGIDAMAGDAINNALSKLGA 403
 Db 361 pmagdtgnglqargagsslgldammagdalinna1gklga 403
 RESULT 4
 ID AAM87639 standard; Protein; 403 AA.
 AAM87639;
 09-MAR-1999 (first entry)
 A hypersensitive response elicitor protein.
 Hypersensitive response elicitor protein; hairpin protein;
 disease resistance; seed quality; insect control; corn borer;
 Lepidoptera larvae; transgenic plant.
 Erwinia amylovora.
 W09854214-A2.
 03-DEC-1998.
 28-MAY-1998; 98WO-US10874.
 30-MAY-1997; 97US-0048109.
 (CORR) CORNELL RES FOUND INC.
 (EDEN-) EDEN BIOSCIENCE CORP.
 Beer SV, Laby RJ, Wei Z;
 WPI: 1999-070210/06.
 N-PSDB; AAV83989.
 New fragments of an Erwinia hypersensitive response elicitor protein
 and related DNA - used to impart disease resistance to plants, to
 increase their growth and to control insects
 Claim 4: Page 10-11; 94pp; English.
 The present sequence represents a hypersensitive response elicitor
 protein (also called hairpin protein) that is able to elicit a
 hypersensitive response in plants. The specification also describes
 hypersensitive response elicitors from other pathogenic organisms.
 The protein, in non-infectious form, is applied to plants to impart
 disease resistance (to a wide range of viral, bacterial and fungal
 pathogens), to improve growth (yield, quantity and quality of seeds,
 to provide earlier germination etc.) and to control insects (e.g. corn
 borers, lepidoptera larvae etc.) The same results are provided by
 transgenic plants expressing the protein.
 Sequence 403 AA:
 Query Match 100.0%; Score 2079; DB 20; Length 403;
 Best Local Similarity 100.0%; Pred. No. 9.6e-156;
 Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 1 MSLNTSGLAGASTWQISIGAGGNNGLGTSRONAGLGNSALGLGGNONDTVNLAGLL 60

|||||
Db 1 msIntsglgaStmqis199ggnngllgterqna199gnal199gnqndvtnqlagll 60
QY 61 TGNMAMMNMGGGGLMGGGGLGNGLGSGGGLGEGLSNALNDMLGSLNTLGSKGNN 120
Db 61 TGNMAMMNMGGGGLMGGGGLGNGLGSGGGLGEGLSNALNDMLGSLNTLGSKGNN 120
QY 121 TTSSTNSPLDQALGINSTSONDSTSGTSDSPDMOOLKMFSEIMQSLFGDCODGT 180
Db 121 tttstnspldqlglnstgndstsgtdstsdspdmqllkmfseimqslfgdcqdgct 180
QY 181 QGSSSGCKOPTBEBQNAKKGVTDALSGLMGNLSQLLNGGLGGGQGGNAGTGLDGSLL 240
Db 181 qgsssgckoptbepqnaaykkyvtcdalslmgnglsqllgng199g99gnaqfclgss1 240
QY 241 GGGGLQNLGSPVYQOLGNNAVGTGIGMKAGTQALNDIGTRHSTSTSFVKKGRAMAKET 300
Db 241 gggglqnlsgpvdyqqlgnavgtg1gmkgag1qalnd1gtrhststsfvkkgramakel 300
QY 301 GQPMDOYPEVFGKPOYOKGPGOEKTDKSWAKALSKPDDGKTPASMEQFNKAKGMIR 360
Db 301 gqfmdypevfgykpgygevtckdkswakalstkpdddgmtpasmeqfnkakgmirk 360
QY 361 PMAGDTGNGMLQARAGAGSSSLGIDAMMAGDAINNMALGKLGAA 403
Db 361 pmagdtgngmlqaragagssslgidammagda1nmalgyklgaa 403
RESULT 5
AA71093
ID AAY71093 standard; Protein; 403 AA.
AC AAY71093;
XX 08-SEP-2000 (first entry)
DE Erwinia amylovora hypersensitive response elicitor #1.
XX Hypersensitive response elicitor; environmental stress resistance;
KM Plant.
OS Erwinia amylovora.
XX WO200028055-A2.
PN 18-MAY-2000.
XX 04-NOV-1999; 99MO-US26039.
XX 05-NOV-1998; 98US-0107243.
XX (EDEN-) EDEN BIOSCIENCE CORP.
XX Wei Z, Schading RL;
PI WPI: 2000-376566/32.
XX N-PSDB; AAD000668.
DR Application of a hypersensitive response elicitor protein to plants to
PT impart stress resistance
XX
XX
PS Disclosure; Page 7-8; 84pp; English.
XX The patent discloses a method to impart stress resistance to plants by
CC applying a hypersensitive response elicitor in a non-infectious form to
CC a plant or seed. The present sequence is a hypersensitive
CC response elicitor protein from Erwinia amylovora. It is
CC is used to impart stress resistance to plants.
XX
XX Sequence 403 AA:
Query Match 100.0%; Score 2079; DB 21; Length 403;

Best Local Similarity 100.0%; Pred. No. 9.6e-156;
Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSINTSGLGASTMQISTIGAGGNNGLGTSRONAGLJGNSALGLGGGNDVYNQAGLL 60
Db 1 msIntsglgaStmqis199ggnngllgterqna199gnal199gnqndvtnqlagll 60
QY 61 TGNMAMMNMGGGGLMGGGGLGNGLGSGGGLGEGLSNALNDMLGSLNTLGSKGNN 120
Db 61 TGNMAMMNMGGGGLMGGGGLGNGLGSGGGLGEGLSNALNDMLGSLNTLGSKGNN 120
QY 121 TTSSTNSPLDQALGINSTSONDSTSGTSDSPDMOOLKMFSEIMQSLFGDCODGT 180
Db 121 tttstnspldqlglnstgndstsgtdstsdspdmqllkmfseimqslfgdcqdgct 180
QY 181 QGSSSGCKOPTBEBQNAKKGVTDALSGLMGNLSQLLNGGLGGGQGGNAGTGLDGSLL 240
Db 181 qgsssgckoptbepqnaaykkyvtcdalslmgnglsqllgng199g99gnaqfclgss1 240
QY 241 GGGGLQNLGSPVYQOLGNNAVGTGIGMKAGTQALNDIGTRHSTSTSFVKKGRAMAKET 300
Db 241 gggglqnlsgpvdyqqlgnavgtg1gmkgag1qalnd1gtrhststsfvkkgramakel 300
QY 301 GQPMDOYPEVFGKPOYOKGPGOEKTDKSWAKALSKPDDGKTPASMEQFNKAKGMIR 360
Db 301 gqfmdypevfgykpgygevtckdkswakalstkpdddgmtpasmeqfnkakgmirk 360
QY 361 PMAGDTGNGMLQARAGAGSSSLGIDAMMAGDAINNMALGKLGAA 403
Db 361 pmagdtgngmlqaragagssslgidammagda1nmalgyklgaa 403
RESULT 6
AA784854
ID AAY784854 standard; Protein; 403 AA.
AC AAY784854;
XX 08-AUG-2000 (first entry)
DE A hypersensitive response elicitor protein.
XX Hypersensitive response; insect control; disease resistance;
KM hypersensitive response elicitor; plant growth; vegetable; crop;
KM ornamental plant.
XX Erwinia amylovora.
XX
XX
FH Key Location/Qualifiers
FT Peptide 169..403 /note="claimed under claim 5"
FT Peptide 210..403 /note="claimed under claim 5"
FT Peptide 267..403 /note="claimed under claim 5"
FT Peptide 343..403 /note="claimed under claim 5"
FT Peptide /note="claimed under claim 5"
FT Peptide 105..179 /note="claimed under claim 7"
FT Peptide 121..150 /note="claimed under claim 7"
FT Peptide 137..166 /note="claimed under claim 7"
FT Peptide 137..156 /note="claimed under claim 7"
FT Peptide /note="claimed under claim 7"
PN WO200020452-A2.
XX 13-APR-2000.
XX 05-OCT-1999; 99MO-US23181.
XX 05-OCT-1998; 98US-0103050.

XX (EDEN-) EDEN BIOSCIENCE CORP.
 PA
 XX
 PI Wei Z, Fan H, Nigemeyer Jr;
 XX WPI, 2000-303745/26.
 DR N-PSDB; AAL4938.
 XX
 PT Hypersensitive response elicitor polypeptides useful for imparting
 PT enhanced growth, disease resistance and insect resistance to plants,
 PT especially vegetables and ornamental flowers -
 PS
 XX Claim 4; Page 8-10; 100pp; English.
 CC The present sequence represents a hypersensitive response elicitor
 CC polypeptide. The specification describes hypersensitive response
 CC elicitor polypeptide fragments, which do not elicit a hypersensitive
 CC response. Instead, the proteins impart disease resistance to plants,
 CC enhance plant growth, and/or control insects. The polypeptide
 CC fragments may be used to these properties to plants. The plants which
 CC may be treated in this way include vegetables, crops and ornamental
 CC plants such as alfalfa, rice, wheat, barley, rye, cotton, sunflower,
 CC peanut, corn, potato, sweet potato, bean, pea, chickory, lettuce,
 CC endive, cabbage, brussel sprout, beet, parsnip, turnip, cauliflower,
 CC broccoli, radish, spinach, onion, garlic, eggplant, pepper, celery,
 CC carrot, squash, pumpkin, zucchini, cucumber, apple, pear, melon,
 CC citrus, strawberry, grape, raspberry, pineapple, soybean, tobacco,
 CC tomato, sorghum or sugarcane, Arabidopsis thaliana, Saltpaulia,
 CC petunia, pelargonium, poinsettia, chrysanthemum, carnation or zinnia.
 XX
 XX Sequence 403 AA:
 SQ
 Query Match 100.0%; Score 2079; DB 21; Length 403;
 Best Local Similarity 100.0%; Pred. No. 9.6e-156;
 Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSLSGSGASTMQLISIGAGGNNGLCTSRONAGLGSNSALGLGGNQNDVTNQLAGIL 60
 Db 1 msltslgastmqslsgagngmlglsrqnaglgsnsalglggnqndvtnqlagl 60
 QY 61 TGMAMMSMNGGGLMGGLGGLGNGLGSGGLGSLGNALNDMLGSLNTLSKGGNN 120
 Db 61 tgmammsmngggglmggglgnglgsgglgslgnalndmlgslntlsgkgnn 120
 QY 121 TTSSTNSPLDQALGINSNDSTSGTSDSPMOQLKMSSEIMQSLFGSGGOGT 180
 Db 121 tstsnspldqlglnstsqndstsgtscdsdpmqqlkmtselmqslfgdgqgdt 180
 QY 181 QGSSSGKQPTGEGEONAKKGVTDALSGLMNGLSQLLNGNGLGAGGOGNAGTGLDSSSL 240
 Db 181 qgsssgkqptegegonakkyvtldalsglmgnglsqllngnglggsggnagtgldgsssl 240
 QY 241 GKGGLQNTSGPVDYQQLGNAVGTEGKAGIOALNDIGTHRSSSTRSFVNKGDRAAKEI 300
 Db 241 gkgglqntsgpvdqyqqlgnavgtegkagiolandigthrsstrsfvnkgdramakei 300
 QY 301 GGFMOQYEVFEGKPOYOKGPGQEVKTDKSNAAKSKPDDOQMTFASMOFNKAKGMIR 360
 Db 301 ggfmdqyevfegkpyokpgqevktddksnwaakskpddqmtfpasmeqfnkakgmir 360
 QY 361 PMAGDTGNGNLQARGAGSSSLGIDAMAGDAINNNALGKLGAA 403
 Db 361 pmagdtgngnlqargagssslgidammagdainnnalgklgaa 403
 RESULT 7
 AAM06598
 ID AAM06598 standard; Protein; 385 AA.
 XX
 AC AAM06598;
 XX
 DT 30-MAR-1997 (first entry)

XX
 DE Hypersensitive response elicitor protein.
 XX
 KW Hypersensitive response; elicitor; Erwinia amylovora; plant;
 KW disease-resistance; Escherichia coli; infiltration; virus;
 KW bacterium; fungus; pathogen; biological control agent.
 XX
 OS Erwinia amylovora.
 XX
 PN W09639802-A1.
 XX
 PD 19-DEC-1996.
 XX
 PF 05-JUN-1996; 96MO-US08819.
 XX
 PR 07-JUN-1995; 95US-0475775.
 XX
 PA (CORR) CORNELL RES FOUND INC.
 XX
 PI Beer SV, Wei Z;
 XX
 DR WPI: 1997-051614/05.
 DR N-PSDB; AAT49314.
 PT
 PT Imparting pathogen resistance to plants - with hypersensitive
 PT response elicitor polypeptide or protein
 XX
 PS Claim 7; Page 46-47; 69pp; English.
 XX
 CC This sequence represents a hypersensitive response elicitor from
 CC Erwinia amylovora, with a mol.wt. of 37,000. The elicitor has a pI
 CC of 4.3, thermostability at 100 deg C for at least 10 min, and
 CC contains no cysteine. The elicitor may be used in a new method for
 CC imparting pathogen resistance to plants, by application of the
 CC elicitor in a non-infectious form to plant cells, by spraying,
 CC infection, leaf abrasion, or plant infection with recombinant
 CC bacteria (non-infectious to the host plant, e.g. Escherichia coli)
 CC expressing the elicitor as a biological control agent, to allow
 CC recombinant protein infiltration into the plant. The method
 CC confers virus, bacterium or fungus disease-resistance on crops and
 CC ornamental plants.
 XX
 XX Sequence 385 AA:
 SQ
 Query Match 92.7%; Score 1928; DB 18; Length 385;
 Best Local Similarity 100.0%; Pred. No. 6.7e-144;
 Matches 372; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSLSGSGASTMQLISIGAGGNNGLCTSRONAGLGSNSALGLGGNQNDVTNQLAGIL 60
 Db 1 msltslgastmqslsgagngmlglsrqnaglgsnsalglggnqndvtnqlagl 60
 QY 61 TGMAMMSMNGGGLMGGLGGLGNGLGSGGLGSLGNALNDMLGSLNTLSKGGNN 120
 Db 61 tgmammsmngggglmggglgnglgsgglgslgnalndmlgslntlsgkgnn 120
 QY 121 TTSSTNSPLDQALGINSNDSTSGTSDSPMOQLKMSSEIMQSLFGSGGOGT 180
 Db 121 tstsnspldqlglnstsqndstsgtscdsdpmqqlkmtselmqslfgdgqgdt 180
 QY 181 QGSSSGKQPTGEGEONAKKGVTDALSGLMNGLSQLLNGNGLGAGGOGNAGTGLDSSSL 240
 Db 181 qgsssgkqptegegonakkyvtldalsglmgnglsqllngnglggsggnagtgldgsssl 240
 QY 241 GKGGLQNTSGPVDYQQLGNAVGTEGKAGIOALNDIGTHRSSSTRSFVNKGDRAAKEI 300
 Db 241 gkgglqntsgpvdqyqqlgnavgtegkagiolandigthrsstrsfvnkgdramakei 300
 QY 301 GGFMOQYEVFEGKPOYOKGPGQEVKTDKSNAAKSKPDDOQMTFASMOFNKAKGMIR 360
 Db 301 ggfmdqyevfegkpyokpgqevktddksnwaakskpddqmtfpasmeqfnkakgmir 360

QY 361 PMAGDTGNGNLQ 372
 Db 361 pmagdtgngnlq 372

RESULT 8

AA045751 standard; Protein; 385 AA.

AA045751;

11-JUL-1994 (first entry)

Erwinia amylovora harpin.

Harpin; hypersensitive response elicitor; HR-elicitor; fire blight;
 Rosaceae; apple; pear; phytopathogenic bacteria; defence reaction;
 hypersensitive reaction and pathogenicity; hrpN; gene cluster; ds.
 Erwinia amylovora.

Key Location/Qualifiers

Misc-difference 372 /note= "corresponds to CAG codon"

Misc-difference 373 /note= "corresponds to CAC codon"

MO9401546-A.

20-JAN-1994.

30-JUN-1993; 93MO-US06243.

01-JUL-1992; 92US-0907935.

(CORR) CORNELL RES FOUND INC.

Bauer DW, Beer SV, Collmer A, He S, Laby R, Wei Z;

WPI; 1994-035054/04.

N-PSDB; AA055751.

Hypersensitive response elicitor protein derived from Erwinia
 amylovora - and DNA encoding it, useful for developing harpin
 inhibitors to prevent e.g. fire blight of fruit

Claim 6; Page 27-28; 47pp; English.

The hrpN gene was isolated from E.amylovora using a 48-fold
 degenerate oligonucleotide probe corresponding to amino acids 9-15
 at the N-terminus of harpin. The 44kd protein encoded by the hrpN
 gene is a hypersensitive response elicitor protein. The harpin is
 thought to be an archetype for HR elicitors from phytopathogenic
 bacteria.

Sequence 385 AA;

Query Match 92.0%; Score 1913; DB 15; Length 385;
 Best Local Similarity 99.5%; Pred. No. 1e-142;

Matches 369; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSINTSGTASTQISIGAGNNGLLGTSRONAGLGNSALGIGGNDVTYNQLAGLL 60

Db 1 msintsgtastqisigagngngllgtscrqnaqlygnsaiglyggnqndvtynqlagll 60

QY 61 TGNMNMNMNMGGGGLMGGLGGLGNGLGSGGGLGEGLSNALNDMLGGSINTLGSKGGNN 120

Db 61 tgnnmnmnmnggggllmggllgngllygsggllgegljnalndmlggsintlgsksqnn 120

QY 121 TTSTTNPRLQALGINTSGNDSTSTSDSSDPMQQLKFKFSIMGSLFDDGDDGT 180

Db 121 ttsttspldqalglintsgndststsdssdpmqqlkfmfseimsglfdgddgt 180

QY 181 QGSSSGKOPTEGEONAKKGVTDALSGLMNGLSQLLGNGGLGGGAGNAGTGLDSSSL 240
 Db 181 qgsssgkoptegeonakkgvtdalsglmnglsqllgngglgggagngagtglgsssl 240
 QY 241 GKGGLONTSGPVDYQOLGNAVGTGIGKAGIOALNDIGTHRSSTRSFVNGGRAMAKET 300
 Db 241 gkgglrntsgpvdqqlygnavgtgigmkagiqalndigthrstsfvngkgramakel 300
 QY 301 GGFMDQYPEVFGKPKQYQKGGGEVKTDDKSMAKALSKPPDDGTFPASMEDCFNAKGMIR 360
 Db 301 ggfmdqy pevfgkpkqyqkgggevktdkswakalskppddgtfpasmegfnkagmlyr 360
 QY 361 PMAGDTGNGNL 371
 Db 361 pmagdtgngnl 371

RESULT 9

AA06597 standard; Protein; 338 AA.

AA06597;

30-MAR-1997 (first entry)

Hypersensitive response elicitor protein.

Hypersensitive response; elicitor; Erwinia chrysanthemi; plant;
 disease-resistance; Escherichia coli; infiltration; virus;
 bacterium; fungus; pathogen; biological control agent.

Erwinia chrysanthemi.

WO9639802-A1.

19-DEC-1996.

05-JUN-1996; 96MO-US08819.

07-JUN-1995; 95US-0475775.

(CORR) CORNELL RES FOUND INC.

Beer SV, Wei Z;

WPI; 1997-051614/05.

N-PSDB; AAT49313.

Imparting pathogen resistance to plants - with hypersensitive
 response elicitor polypeptide or protein

Claim 4; Page 44; 69pp; English.

This sequence represents a hypersensitive response elicitor from
 Erwinia chrysanthemi, with a mol.wt. of 34,000. The elicitor is
 thermostable, has a glycine content of over 16%, and has no cysteine.
 The elicitor may be used in a new method for imparting pathogen
 resistance to plants, by application of the elicitor in a non-
 infectious form to plant cells, by spraying, injection, leaf
 abrasion, or plant infection with recombinant bacteria (non-
 infectious to the host plant, e.g. Escherichia coli) expressing the
 elicitor as a biological control agent, to allow recombinant
 protein infiltration into the plant. The method confers virus,
 bacterium or fungus disease-resistance on crops and ornamental
 plants.

Sequence 338 AA;

Query Match 34.6%; Score 718.5; DB 18; Length 338;

Best Local Similarity 42.8%; Pred. No. 6.9e-49;
 Matches 173; Conservative 41; Mismatches 111; Indels 79; Gaps 11;

QY	13	MOIST-GGAGGNNGELTGRSRNAGIG-NSA----	LGLGGGNQNDTVNOJLGTGMMAM	67
Dd	1	mgttkahlgddldvsglgaq--glkglnsaa	sllgssvdsktsldktltalsmm---	55
QY	68	SMMGGGGLMGGLGGGGLGGGGLGGSGGLE	EGLSNALMDMLGSSLMTLTKSSKGGNNFTSTTN	127
Dd	56	-----fggalagylgas-skglgmsnqlgs	fgn-----gagasnllypk-	96
QY	128	PUDQALGINSTRSQNDNSTISGDTSDSSD	PMQQLMKPSELMQSLFC-----DGODG	177
Dd	97	-----sggdals-----kmfdkalddlgh	dvtkltnsqnq	120
QY	180	TQGSSSGCKOPTGEGBONAYKKGYVDDAL	SLGMNGHLSOLTLGGKGLGGGGGMAAGLDGSS	230
Dd	129	lanmhnasgmttgmnaafgsgymalsslign	lgyqsm-----sfagsp	17
QY	240	LGGKGLQMLSGRPVDYQQLGNAVGTGIGR	KACIQALINDIGHRSSTRFVNKGDRAMAKE	299
Dd	175	lgaagllglsagaflnglnaigmvgqnaa	lsalnvsrhvdgnrhlfvdkedrgmae	23
QY	300	IGOFNDQPEVFEGKPRQYKCGEQEVKTPD	KSWAKALSRPDDGHTPRASMEFPNAKGHIT	355
Dd	235	lqglmdqypelfgkpxekykdswspktdk	kawakalskpdddgmtgasmokfrqamgmk	29
QY	360	RPMAGDTGNGMLQARAGAGSSGLIDPAMA	GCAINNMALGKICAA	403
Dd	295	savgdfctgnlnhlrgaggaalsgldaav	gvdkianmslglkana	338
RESULT	10			
AAM87638	ID	AAM87638 standard; Protein: 338 AA.		
XX	AC	AAM87638;		
XX	DT	09-MAR-1999 (first entry)		
XX	DE	A hypersensitive response elicitor protein.		
XX	KW	Hypersensitive response elicitor protein; hairpin protein; disease resistance; seed quality; insect control; corn borer; Lepidoptera larvae; transgenic plant.		
XX	OS	Erwinia chrysanthemi.		
XX	PN	WO9854214-A2.		
XX	PD	03-DEC-1998.		
XX	PF	28-MAY-1998; 98WO-US10874.		
XX	PR	30-MAY-1997; 97US-0048109.		
XX	PA	(CORR.) CORNELL RES FOUND INC.		
XX	PA	(EDEN-) EDEN BIOSCIENCE CORP.		
XX	P1	Beer SV, Laby RJ, Wei Z;		
XX	PT	WIPI; 1999-070210/06.		
XX	DR	N-PSDB; AAV83988.		
XX	PT	New fragments of an Erwinia hypersensitive response elicitor protein and related DNA - used to impart disease resistance to plants, to increase their growth and to control insects		
XX	PS	Disclosure; Page 7-8; 94pp; English.		
CC	XX	The present sequence represents a hypersensitive response elicitor protein (also called hairpin protein) that is able to elicit a hypersensitive response in plants. The specification also describes hypersensitive response elicitors from other pathogenic organisms. The protein, in non-infectious form, is applied to plants to impart		

CC disease resistance (to a wide range of viral, bacterial and fungal pathogens), to improve growth (yield, quantity and quality of seeds, to provide earlier germination etc.) and to control insects (e.g. corn borers, Lepidoptera larvae etc.) The same results are provided by CC transgenic plants expressing the protein.

xx
SQ Sequence 338 AA;

Query Match 34.6%; Score 718.5; DB 20; Length 338;
Best Local Similarity 42.8%; Pred. No. 6.9e-49;
Matches 173; Conservative 41; Mismatches 111; Indels 79; Gaps 11:

QY 13 MOIST-GGAGCGNGLIGTSRQNAIGGG-NSA---LGISGGKNDPYNVNLALITGMAMMM 67
||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 1 ngitlkahlgdlygslygaq--gikjinsaaaslyssvdklsstldkltsltsmm--- 55

QY 68 SMMGGGGLMGGLGGLGGLGSLGEGSLNGLBGLSNALNDMLGSGLSLNTFGSKGNNSTTNS 127
56 -----fggalagglgas-skylgmngllygstgn-----gaqgasnllsyvk- 96
||| ||| | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

QY 128 PLDQALINSTSQNDSTGSTDTSDSSDPMQQLKMFSEIMQSIFG-----DGDDG 179
Db 97 -----sgdals-----kmfdkalddllghndtyckllengsq 128
||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

QY 180 TGGSSSGCKOPTPEGRONAYKKGVTDALSGLMNGISQLIGNGGLGGGCGGNAGTLDDSS 239
Db 129 Iansmlnasqmqlgymnaifsgysnalaisllnglyqsm-----sgfsqps 174
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

QY 240 LGKGKLONLSCGPVDYQOOLGNNAVGTGICMKAGIQIALNDIGTHRHSSTRSPVNKGDRAAKE 299
Db 175 lgaagllgllsgagafrnqalgnaligmvgvqnuaalsalnvtshvdgmrtfvdkedrgmake 234
||| ||| ||| : ||||||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

QY 300 IOGFMDQYPEVEGKPOYOXKPGQEYKTTDKSMAKALSKEPDGDMTPASMEOPNKAKGMIR 359
Db 235 lqgfmdqypeilfgkyeqdydwgswspktiddkswakalskpdddgmtygasndkfrtgamgmk 294
||||||| : ||||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

QY 360 RMAEDTGNGNTQARGAGSSLCIDPAAAGAIDLINNMALGKLCGA 403
Db 295 savagdtgtntlnlrgaagasigidaavvydklanmslqklana 338
: ||||| : ||||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

RESULT 11
AAW82407
ID AAW82407 standard; Protein; 338 AA.
XX AC
XX AAW82407;
XX DT 23-FEB-1999 (first entry)
XX DE
XX E. chrysanthemi HrpN-ECH protein.
XX KM Hypersensitive response elicitor protein; hrpN-Ech; pathogen resistance;
XX plact; transformation; pathogen-inducible promoter.
XX Erwinia chrysanthemi.
OS
PN US5850015-A.
XX PD 15-DEC-1998.
XX PF 07-JUN-1995; 95US-0484358.
XX PR 07-JUN-1995; 95US-0484358.
PA (CORR) CORNELL RES FOUND INC.
XX PI Bauer D, Collier A;
XX DR WPI: 1999-069852/06.
DR N-PsDB: AAV73494, AAV73507.
XX PT DNA encoding Erwinia chrysanthemi hypersensitive response protein

PT hrpn - useful for imparting pathogen resistance to plants
XX
PS Claim 2; Column 29-30; 27pp; English.
v

CC This sequence represents a novel *Erwinia chrysanthemi* protein, hrpN-Ech,
CC that elicits a hypersensitive response in plants. The encoding DNA can be
CC used for imparting pathogen resistance to plants, by transforming a
CC plant with a vector containing the DNA and a pathogen-inducible promoter.
XX
XX
SQ Sequence 338 AA;

SQ Sequence 338 AA;

	Query Match	34.6%; Score 718.5; DB 20; Length 338;
	Best Local Similarity	42.8%; Pred. No. 6.9e-49;
	Matches 173; Conservative	41; Mismatches 111; Indels 79; Gaps 11.
OY	13 MOISI-GGAGNNGLLGTSRONAGIG--NSA---LALGGNONDYNOLAGLTGMMMA	67
Db	1 mqltkahlgdglgyssgiyaq-gltkylnsaasslgyssvdklsstfkltaalmsm---	55
OY	68 SMMGGGMLGGGGICGCGCNGLGSGCGLEBGLSNALNDMLGSLNLTGSNGNNSTTYS	127
Db	56 -----fggalaqglgas-skljgmngnlyqgsfgn-----gaqgasnlisvpk-	96
OY	128 PLDDALGINSTRSOUNDSTGDTSDSSDPMOQLKMFSEIMQSLFG-----DGQQP	179
Db	97 -----sggdals-----kmfdakddllghndvtckltnsq	128
OY	180 TQGSSSGCKRPEEGEQNAVKKGVTDALSGLMGNLSQLGNGLCGGCGGAGATGLDSS	239
Db	129 laanmlasqmtctgumnaifgsvnnalssllygnlyqsm-----sfjsqps	174
OY	240 LGGKGILMTSGPVDRVOOAGNAVGTGIGKACGIALQANDIGTHRHSSRPVNKGDMARKE	299
Db	175 lgaagliglisagaafnqilgnaigmvgvqnaaisalstvnsthvdgmnrhlvtvdfedymate	234
OY	300 IGEPMDOYPEVFGRFOYKGPQEVEKTDDSKWAKLLSPRDDGMPPASMEQNPKAKGIT	359
Db	235 lggfmfygpelifykpeykdgwspsktiddkswakelskpdddmgyamsgndkfrtgmgmk	294
OY	360 RPMAGDTGNGNLQARGAGSSSLGIDAMMAGDALNNMALGKLCA	403
Db	295 savagdtegnclnlrtgagyasglidaavvgldlamsltyklana	338

RESULT 12

AAV71092
ID AAV71092 standard; Protein; 338 AA.

AC AAY71092;

DT 08-SEP-2000 (first entry)

DE *Erwinia chrysanthemi* hypersensitive response elicitor.

KM Hypersensitive response elicitor; environmental stress resistance,
KM plant.

OS Erwinia chrysanthemi.

Key	Location/Qualifiers
EH	19
FT	Misc-difference
FT	Encoded by GGGCTGGCT
FT	/note-

PN W0200028055-A2.

PD 18-MAY-2000.

PF 04-NOV-1999; 99WO-US26039.

PR 05-NOV-1998; 98US-0107243.

PA (EDEN-) EDEN BIOSCIENCE CORP.

XX Wei Z, Schading RL;

DR WPI; 2000-376566/32.

XX

PT Impart stress resistance -

PS Disclosure; Page 4-5; 84pp; English

CC stable and to impart stress resistance to plants.

SQ **Sequence** **338 AA;**

Query Match	34.6%	Score 718.5;	DB 21;	Length 338;
Best Local Similarity	42.8%	Pred. No. 6.9e-49;		
Matches 173; Conservative	41;	Mismatches 111;	Indels 79;	Gaps 11.

RESULT 13

AA184853
ID AY84853 standard; Protein; 338 AA.
yy

AC. AA84853;

DT 08-AUG-2000 (first entry)

DE A hypersensitive response elicitor protein.

KW Hypersensitive response; insect control; disease resistance;
KW hypersensitive response elicitor; plant growth; vegetable; crop
KW ornamental plant.

OS *Erwinia chrysanthemi*.

EH	Key	Location/Qualifiers
FT	Misc-difference	20
FT		/note- "encoded by CTGGCTGCT"

X)

RESULT 2

D70807

hypothetical glycine-rich protein RV3514 - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000

C:Accession: D70807

R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

R:Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

R:Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987

A:Accession: D70807

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 11489 <COL>

A:Cross-references: GB:AL022022; GB:AL123456; NID:g3261554; PIDN:CAA17751.1; PID:g292445

A:Experimental source: strain H37RV

C:Genetics:

C:Superfamily: collagen alpha 1(IV) chain

Query Match

12.7%; Score 264.5; DB 2; Length 1489;

Best Local Similarity 26.1%; Pred. No. 4.3e-09;

Matches 116; Conservative 27; Mismatches 149; Indels 153; Gaps 16;

QY 3 LNTSGLGASTMQT-----SIGAGGNNGI-----LGTSRON 33

DB 955 LMTDGLSSATSGTGTGTGKGCTGAGDPSAGCTGTGTGAGNAGAGIANTGTAGN 1014

QY 34 AGAGNSALGIGGNDNTYNOL-----AGLTGMMMMMMGGGGLMGGLGGG 87

DB 1015 AGTGGDGGG-GNGGGDSSSGIGGPGFAGAGCGAGCGAGSSAGGTNGSGGAGAG-G 1072

QY 88 LGGSGGLGEGLSNALNDMLGGS--LNTLGSKGKNTTSTNSPLDQALGINSTQNDST 145

DB 1073 GCGAGTAGISFNSNSGCTGTGTGCGTGGGCGN-----AGTGADGPRGKT 1119

QY 146 SCNDTSDSDPMQQLKMESEIMQSLFGDQDGTGSS--SGKQPIEGEDNATKKVTTD 204

DB 1120 GGTGGTGGG--GAGSSGAGNFGCTGTGTGCTGCTGCKGCMG 1159

QY 205 ALSGMNGLSQLLGNGLG-----GGGGNAGTGLDSSSLGKGLNLSPVYDQ 256

DB 1160 GTAGDPPGCGD--GNAGVGKGTGTGNGSGGTGTGTGAGAGAGAGLANTGTA--- 1213

QY 257 LGNAVGTGIMKAGIQAINDIGTHRSSTRSFVNKGDRAKKEIGOFMDQYPEVFKRPOY 316

DB 1214 -GNA---GIG-----GDGGGGGNGGQ--GDSSSLGCGPGF 1243

QY 317 OKGGGQEVKTDKSKAKALSKPDDGMPASMEQFNKAKIKRPMAGDTGNGMLQARGA 376

DB 1244 AGGCGG-----KKGAGGAGNAGTGTGNGSGA 1267

QY 377 GGSSLDGIDAMMAGDAINMALGKLG 401

DB 1268 GGAGGGGAGAGAGISFNSGNGGTG 1292

RESULT 3

hypothetical glycine-rich protein RV0297 - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 21-Jan-2000

C:Accession: B70523

R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

R:Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

R:Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987

A:Accession: B70523

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-591 <COL>

A:Cross-references: GB:296800; GB:AL123456; NID:g3261800; PIDN:CA809596.1; PID:e32165

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: RV0297

C:Superfamily: unassigned collagens

Query Match

12.3%; Score 255; DB 2; Length 591;

Best Local Similarity 26.8%; Pred. No. 5.8e-09;

Matches 123; Conservative 24; Mismatches 168; Indels 144; Gaps 20;

QY 7 GLGASTMQISTGAGNNGLLGTSRONAGLGGGNDNTYNOLAGLTGMMMM 66

DB 141 GNGSGAPGAGAGAGAGAAFGF-----NCGNCGDAGAGANG-----AGTAGWFFG 187

QY 67 MSMMGGGGLMG-----GLGGGGLGEGLSNALNDML----- 106

DB 188 FGNGGAGGIGVAGINGLGGAGDGGNAGFEFGNGGNGMGAGACVMAVNPGLATPVT 247

QY 107 -----GSLNTLGSKG--GNNTTSTNSPLDQALGINSTQNDSTSG-----TDSTSDS 154

DB 248 PAANGGNCINLVGPYRAGGADGANGSAICQAGAGDGNMSTSGIGTGAOTGGAGA 307

QY 155 SDPMQQLKMESEIMQSLFGDQDGTGSSSGKQPIEGEDNATKKVTTALSGLMGNTL 214

DB 308 G-----GAGGDPAGCGNCGSVET-----GATGS--SASGNGA 342

QY 215 SGLLNGGIG-----GGGGNAGTGLDSSSLGKGLNLSPVYDQ 252

DB 343 TG--GNGVGAPGAGAGNGGVSCTVNTAGAGKGGNGGAGG--GGGGSVLSLV 399

QY 253 DYQQLGNNAVGTGIMKAGIQAINDIGTHRSSTRS-----FVNKGDRAKKEIGOFMDQY 308

DB 400 GDSNGGAGCGD--GAGVSATDLAGTGTGRCGNGHGIMTGNCGDGGAGGAGV----- 449

QY 309 EVFGKPOYQKPGQEVKTDKSKAKALSKPDDGMPASMEQFNKAKIKRPM-----AG 364

DB 450 -----GVG-----GAGAAAGTGGHGGD-----GSVNTPTGSGEAG 480

QY 365 DTGNGNLQARAGAGSSIGIDAMMAGDAINMALGKLGAA 403

DB 481 DGGKGLGCGDGGGNGTIG--OFGAGAGAGAGAGGAGAGA 517

RESULT 4

hypothetical glycine-rich protein RV3388 - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000

C:Accession: C70974

R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

R:Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

R:Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987

A:Accession: C70974

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-731 <COL>

A:Cross-references: GB:AL009198; GB:AL123456; NID:g3242262; PIDN:CAA15773.1; PID:g266

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: RV3388

C:Superfamily: elastin

Query Match 12.0% Score 249; DB 2; Length 731;
 Best Local Similarity 28.7% Pred. No. 1.8e-08;
 Matches 96; Conservative 23; Mismatches 101; Indels 114; Gaps 15;

QY 2 SLNTSGAGTMOISIGAGNNGILGTSRQ-----NAGLGN-SALGLG 45
 Db 402 SAGAGGAGAS-----GGTGNAGLIGAGHGAGAGAGAGNQTGCGNAGAGAGAGAG 455
 QY 46 G---GNNDVYNQAGLITGMMMSMMGGGGLMGGLG-GGLG-----NLGSSGG 93
 Db 456 GOLYNGGDDGNGAG-----GANIANGNGSPGCAAGHAGAGAGSARLIGAGHGCGG 508
 QY 94 LGEGLSNALNDMLGSLTGLSKGN-----NTSTNSPLD 130
 Db 509 AGGTWARRADALIGT-----GGDGNNGNGLISGNAGAGHGAGSSSTATTTGTPT 564
 QY 131 QALGINSTQNDSTSGTSDSDSPMQLKMFSEIMOSLFGDGDGTGSSSGGKOP 190
 Db 565 GATGNGNGAGAGTGTGSG-----GIGGNGAGGTGGMAG----- 602
 QY 191 TEGEONAYKKVTDAL-----SGLMGNLSOLLNGGLG--GGCGNAGTGLDSSSLGCK 243
 Db 603 -----VALSVSTGGLGNGSGGGLFGGCGGLFNGGAGAGVATGNGSGIGPASYG 657
 QY 244 G-----LNLGSPVDYQOLGNNAVGTGIMKAG 270
 Db 658 GKGAGVAGAGLAG-----QIGNGSGSGSGGAGG 686

RESULT 5

hypothetical glycine-rich protein RV1068c - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
 C:Accession: B70893
 R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellom, S.; Squares, S.; Nature 393, 537-544, 1998
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A:Reference number: A70500; MUID:98295987
 A:Accession: B70893
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-463 <COL>
 A:Cross-references: GB:AL021897; GB:AL123456; NID:g3256022; PIDN:CAA17184.1; PID:g289670
 A:Experimental source: strain H37RV
 C:Genetics:
 A:Gene: RV1068c
 C:Superfamily: Phaeococcus glycine-rich cell wall protein 1.8

Query Match 11.9% Score 248; DB 2; Length 463;
 Best Local Similarity 31.8% Pred. No. 1.2e-08;
 Matches 93; Conservative 17; Mismatches 112; Indels 70; Gaps 13;

QY 8 LGASTMOISIGAGNNGILGTSRONAGLIGNSALGLGNGNDVYNQAGLITGMMMM 67
 Db 203 IGAPVAGAGAGAGTATLFG-----NGAGAGAGAGAGAGAGAGAGAGAGAGAG 256
 QY 68 SMGGGGLMGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLG 127
 Db 257 GTGGGGGNAAGNGAG 301
 QY 128 PLDPAALGINSTQNDSTSGTSDSDSPMQLKMFSEIMOSLFGDGDGTGSS--- 184
 Db 302 GIDPG---NGCGAGTGNAGNGHGGSA-----AKLFGDAGAGAGAGAGAGAG 344
 QY 185 -----SGCKQPTTEGONAYKKVTDALSGLMGNLSOLL-----GNGGLGAGG 229
 Db 345 GGTGGGGGFGGGTGTGNGNGNAGHAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 404

QY 230 NAGTGLDSSIGCKGLQNLGSPVDYQOLGNNAVGTGIMKAG-----IQALNDIG 278
 Db 405 NCGNGD-----GCKG-----GDAQLIGNGNGNGGKGGTGLMPGINGTG 445

RESULT 6

hypothetical glycine-rich protein RV3507 - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
 C:Accession: E70806
 R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellom, S.; Squares, S.; Nature 393, 537-544, 1998
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A:Reference number: A70500; MUID:98295987
 A:Accession: E70806
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1381 <COL>
 A:Cross-references: GB:AL022022; GB:AL123456; NID:g3261554; PIDN:CAA17744.1; PID:g2922
 A:Experimental source: strain H37RV
 C:Genetics:
 A:Gene: RV3507
 C:Superfamily: collagen alpha 1(IV) chain

Query Match 11.9% Score 248; DB 2; Length 1381;
 Best Local Similarity 26.8% Pred. No. 4.3e-08;
 Matches 117; Conservative 24; Mismatches 180; Indels 116; Gaps 16;

QY 7 GIGASTMOISIGAGNNGILGTSRONAGLIGNSALGLGNGNDVYNQAGLITGMMMM 66
 Db 822 GIGAGTTTATVNGNGMNG--DGGNGMAGAGAGAGAGAGAGAGAGAGAGAGAG 864
 QY 67 MSMMGGGLMGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLG 126
 Db 865 -SGAGSGGNGAG 919
 QY 127 SPLDPAALGINSTQNDSTSGTSDSDSP--MQLKMFSEIMOSLFGDGDGTGSS 184
 Db 920 ACPVATSLPTPNGAG 979
 QY 185 SGCK-----QTEGEONAYKKVTDALSGLMGNLSOLLNGGLG----- 224
 Db 980 GGDGADAVNAVVPTEPQ-----AATGTAGSAGDPTGNGGPGTGPSPVAPPP 1029
 QY 225 -----GGGAG 267
 Db 1030 TPTVOAGGAG 1080
 QY 268 KAGIQAALNDIGTHHSSTSPFNKCDRAKATGPMQYPEVFGKPYQKPGCEVKT 327
 Db 1081 -----GNASATGTGNVANAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1119
 QY 328 DKSNKALSKPPDDGTPASMEQFNKAKMIRPMAGTGTGNGNLARAGAGAGAGAG 387
 Db 1120 DGSAGTAGNGNG-----NATNGT-----AGPAGAGAGAGAGAGAGAGAG 1166
 QY 388 AGDAIN--NMAIGKLA 402
 Db 1167 TGTAGNGNGAG 1183

RESULT 7

hypothetical glycine-rich protein RV3367 - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 18-Feb-2000
 C:Accession: F70971
 R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon

[illegible]

RESULT 10
 A70869
 hypothetical glycine-rich protein Rv2490c - Mycobacterium tuberculosis (strain H37Rv)
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence,revision 17-Jul-1998 #text,change 20-Jun-2000
 C:Accession: A70869
 R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feldwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A:Reference number: A70500; MUID:98295987
 A:Accession: A70869
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1660 <COL>
 A:Cross-references: GB:AL021246; GB:AL123456; NID:g3261507; PIDN:CAA16067.L; PID:g279155
 A:Experimental source: strain H37Rv
 C:Genetics:
 A:Gene: Rv2490c
 C:Superfamily: collagen alpha 1(IV) chain

[illegible]

QY 292 GDRAAKAEIGGEMDQPEVFGKPPQXOKPEQGEVTDDESKAKALSKRFDDDDMPASMEQF 351

Db 368 GSAGLLGVYGRAGDGGAGGGGGISGAPEDG- GAAGNGSGWTAA-----GDGG----- 413

QY 352 NKAKCMIRPMAGDTGNGCNLQARAGSGSSLTIDMAMGDAIINNALKRLG 401

Db 414 --AGG-----HGGDPGLGG--AGGAGGASGGAGAG-RAQ--ANGIAGAGNDG 451

RESULT 11
 E70820
 hypothetical glycine-rich protein Rv0977 - *Mycobacterium tuberculosis* (strain H37Rv)
 C:Species: *Mycobacterium tuberculosis*
 C:Date: 17-Jul-1998 #sequence.revision 17-Jul-1998 #text.change 20-Jun-2000
 C:Accession: E70820
 R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
 ; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete geno
 A:Reference number: A70500; MUID:98295987
 A:Accession: E70820
 A:Status: Preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-923 <COL>
 A:Cross-references: GB:AL021999; GB:AL123456; NID:g3261538; PIDN:CAA17576.1; PID:g2921
 A:Experimental source: strain H37Rv
 C:Genetics:
 A:Gene: Rv0977
 C:Superfamily: elastin

[illegible]

DB 257 NPSGAADGAGADA-----RLINGGDDG---GSVCAAPTG--- 289
QY 196 NAYKRGVTDALSGLMGNLSQLLNGSLG---GCGNAGTGLDSSSLGCKLONLSGPV 252
DB 290 -----IGN------GNGGNGMWLYGDDGSGSTLQDFSDGGTG----- 322
QY 253 DYQOLGNAVGTGIGKACIQAALNDIGTRHSSTRSFVANKGRAMAKEIGQFMDQYFEVFG 312
DB 323 -----GNMGNGDGGNGGFFPDGNGDGGTGCTILINGGD----- 358
QY 313 KPQYQKPGQEVKTDKSWAKALSKPDDGQMTPASMEQFNKAKGMRPMAGDTGN--- 368
DB 359 -----GNSVQTD-----GFL-RHGHDGCGNAVGL 382
QY 369 -GNLQARGGSSSLGIDAMMAGDAINN---MALCKLGA 402
DB 383 IGNCGAGAGSGNCTGFAPGGSGGNGGNCALLVGNGGA 421

RESULT 15

B70812
hypothetical glycine-rich protein Rv0834c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: B70812
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987
A:Accession: B70812
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-882 <CDL>
A:Cross-references: GB:AL022004; GB:AL123456; NID:g3261550; PIDN:CAAI7640.1; PID:g291689
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: Rv0834c
C:Superfamily: elastin

Query Match 11.5%; Score 238.5; DB 2; Length 882;
Best Local Similarity 27.2%; Pred. No. 1e-07;
Matches 113; Conservative 36; Mismatches 128; Indels 139; Gaps 23;

QY 7 GLGASTMOISIGAGGNN-GLIGTSRONAGLGNLSALIG-----GQNNDTVNQLAGLL 60
DB 521 GGGASVGTGGSGAGGDDGFVGA-----GNGGNAGTGVGVANGCN-----GSSA 568
QY 61 TGMAMMMMSMGGGGLMGGLGGLGGLGSGGLGGLSNAALN--DMLGSLNT-IGSKG 117
DB 569 TC---ALAAVCGAGAGAGATGCTG-GEFGAGGAGSARGLIFALGAGAGAGDASTGVGCPG 624
QY 118 GNNSTSTNSPLDQALGINST-SQNDSTSTGTDSTSDSSDPMOQLKMFSEIMQSLFQDG 176
DB 625 GGGGTGTASSPFGIAIAGGAGAGAGATGATGGAGDGFEGT-----AVLGLQFG-G 678
QY 177 ODGTGSSSGGKQPTGEGQNAKKGVTDLASL--MGNGLSOLLG-----NGGLGGGCGGN 230
DB 679 AAGAGAGANT-----DGAITGAGGFGGAGAGIAFLGFSVYLHGAGGAGGTA 725
QY 231 AGTGLDSSLGSKGLQNLGSPVDYQOLGNAVGTGIMKAG-----IQALNDIGTRHSST 285
DB 726 TGTGNGGAGGCGG---LSSPV-----ILGIGAGAGCGDGGALGVLCMGCG----- 769
QY 286 RSFVKKGDAMAKEIQFMDQYFEVGRKFOYOKPGQEVKTDKSWAKALSKPDDGQMT 345
DB 770 ---DGDGGEVAVGIAG-----GAG----- 788
QY 346 ASMEQFNKAKGMRPMAGDTGNLQARGAGSSSLGIDAMMAGDAINNMLGKLG 401

DB 789 -----GAGC-----AAPTGNG-----GAGCN-----GSDALGLVGVGNG 818

Search completed: October 25, 2001, 11:44:38
Job time: 74 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 25, 2001, 11:44:19 ; Search time 13.1 Seconds
(without alignments)
1053.814 Million cell updates/sec

Title: US-09-412-100-23

Sequence: 1 MSINISGLASTMQLISIGGA.....DAMAGDAINNALGKLGAA 403

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues
Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2050	98.6	403	1 HRPN_ERMAN	001099 erwina amy
2	828.5	39.9	356	1 HRPN_ERMAN	047279 erwina car
3	718.5	34.6	340	1 HRPN_ERMAN	047278 erwina chr
4	248	11.9	463	1 YAB6_MYCTU	053416 mycobacteri
5	242	11.6	778	1 Y034_MYCTU	P71933 mycobacteri
6	241.5	11.6	957	1 Y278_MYCTU	P56877 mycobacteri
7	238.5	11.5	515	1 Y140_MYCTU	050594 mycobacteri
8	238	11.4	1901	1 Y208_MYCTU	053553 mycobacteri
9	235.5	11.3	914	1 WA22_MYCTU	006794 mycobacteri
10	234.5	11.3	543	1 YP91_MYCTU	050630 mycobacteri
11	233.5	11.2	338	1 GRP_ARATH	P27483 arabidopsis
12	233.5	11.2	434	1 YK98_MYCTU	010707 mycobacteri
13	232	11.2	801	1 Y747_MYCTU	053810 mycobacteri
14	229.5	11.0	603	1 YD25_MYCTU	010637 mycobacteri
15	214.5	10.3	384	1 GRP1_PERTX	P09789 petunia hyb
16	213.5	10.3	481	1 LORI_MOUSE	P18165 mus musculu
17	211.5	10.2	747	1 SPDI_MOUSE	P18837 nephila cla
18	210	10.1	1156	1 G1H4_CAEBL	076743 caenorhabdi
19	202.5	9.7	498	1 Y118_MYCTU	056615 mycobacteri
20	199	9.6	937	1 HRY1_CANAL	P46591 candida alb
21	196.5	9.5	465	1 GRP2_PHAVU	P10496 phaseolus v
22	193	9.3	183	1 GRP2_ORISA	P23834 oryza sativ
23	191	9.2	252	1 GRP1_PHAVU	P10495 phaseolus v
24	191	9.2	316	1 LORI_HUMAN	P23490 homo sapien
25	190	9.1	641	1 EBNI_EBV	P03211 Epstein-Bar
26	186.5	9.0	1113	1 N116_YEAST	002630 saccharomyc
27	186	8.9	749	1 TROP_HUMAN	Q12816 homo sapien
28	184.5	8.9	419	1 CSP_PLACM	P08676 plasmodium
29	178.5	8.6	959	1 N100_YEAST	002629 saccharomyc
30	177.5	8.5	1224	1 PER_DROME	P07663 drosophila
31	174.5	8.4	672	1 PHX5_MOUSE	P08399 mus musculu
32	174	8.4	212	1 BEG1_SCHYA	P19470 schistosoma
33	174	8.4	700	1 N0NA_DROME	Q04047 drosophila

34	173.5	8.3	1380	1 DDX9_MOUSE	070133 mus musculu
35	172	8.3	5263	1 FBOH_BOMO	P05790 bombyx mori
36	171.5	8.2	1208	1 PER_DROTA	Q24767 drosophila
37	169	8.1	526	1 FUS_HUMAN	P35637 homo sapien
38	169	8.1	1596	1 MAM_DROME	P21519 drosophila
39	167	8.0	401	1 CSP_PLACG	P08674 plasmodium
40	166.5	8.0	393	1 CSP_PLACB	P14593 plasmodium
41	165.5	8.0	207	1 BGG2_SCHYA	P19469 schistosoma
42	164.5	7.9	404	1 CAZ_DROME	Q27294 drosophila
43	164	7.9	518	1 FUS_MOUSE	P35635 mus musculu
44	164	7.9	622	1 RICI_HUMAN	P35527 homo sapien
45	160	7.7	182	1 KZC3_BOVIN	P04261 bos taurus

ALIGNMENTS

RESULT 1

ID	HRPN_ERMAN	STANDARD:	PRT:	403 AA.
AC	001099;			
DT	01-APR-1993 (Rel. 25, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DT	30-MAY-2000 (Rel. 39, Last annotation update)			
DE	HARPIN (HARPIN-ER).			
GN	HRPN.			
OS	Erwinia amylovora.			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OX	Erwinia.			
NC	NCBI_TaxID=552;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 1-15.			
RC	STRAIN-EA321;			
RX	MEDLINE-92320301; PubMed-1621099;			
RA	Wei Z.-M., Laby R.J., Zumoff C.H., Bauer D.W., He S.Y., Collmer A.,			
RA	Beer S.V.;			
RT	"Harpin, elicitor of the hypersensitive response produced by the			
RT	plant pathogen Erwinia amylovora.";			
RL	Science 257:85-86(1992).			
RN	[2]			
RP	REVIEWS.			
RC	STRAIN-EA321;			
RA	Laby R.J., Kim J.F., Beer S.V.;			
CC	Submitted (MAR-1999) to the EMBL/Genbank/DBJ databases.			
CC	-1- FUNCTION: ELICITS THE HYPERSENSITIVE RESPONSE (HR) IN THE PLANT			
CC	UPON INFECTION. HARPIN ELICITS HR IN NON-HOSTS AND IS ALSO			
CC	REQUIRED FOR PATHOGENICITY IN HOST PLANTS.			
CC	-1- SUBCELLULAR LOCATION: SECRETED; VIA THE HRP SECRETION PATHWAY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement. (See http://www.isb-sdb.ch/announce/			
CC	or send an email to license@sdb.ch).			
CC	-----			
DR	EMBL; M92994; AAC31644.2; -			
KM	Hypersensitive response.			
FT	DOMAIN 1 270			
FT	DOMAIN 63 70			
FT	POLY-MET.			
FT	SEQUENCE 403 AA; 39697 MW; 146FA62351DBE87 CRC64;			

Query Match 98.6%; Score 2050; DB 1; Length 403;

Best Local Similarity 99.0%; Pred. No. 4.1e-112; Mismatches 4; Indels 0; Gaps 0;

QY	1 MSINISGLASTMQLISIGGAGNNGILGTSRONAGLGNSALGGLGGGNDVYNQLAGLL 60
DB	1 MSINISGLASTMQLISIGGAGNNGILGTSRONAGLGNSALGGLGGGNDVYNQLAGLL 60
QY	61 TGMNMMMSMAGGGGGLMGGLGGGLGGGLGGSGGLGEGCLSNALNDMLGGSINTLGSGGNN 120

[illegible]

RESULT	2			
HRPN_ERMCA	STANDARD;	PRT;	356 AA.	
ID	HRPN_ERMCA			
AC	Q47279;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	15-DEC-1998 (Rel. 37, Last annotation update)			
DE	HARPIN (HARPIN-ECC) (FRAGMENT).			
GN	HRPN.			
OS	Erwinia carotovora.			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OC	Pectobacterium.			
OX	NCBI_TaxID=554;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=71;			
RX	MEDLINE=96405946; PubMed=8810071;			
RA	Cul Y., Madl L., Mukherjee A., Dumenyo C.K., Chatterjee A.K.;			
RT	"The RsmA-mutants of Erwinia carotovora subsp. carotovora strain			
RT	Ecc71 overexpress hrpNcc and elicit a hypersensitive reaction-like			
RT	response in tobacco leaves.";			
RL	Mol. Plant Microbe Interact. 9:565-573(1996).			
CC	-1- FUNCTION: ELICITS THE HYPERSENSITIVE RESPONSE (HR) IN THE PLANT			
CC	UPON INFECTION. HARPIN ELICITS HR IN NON-HOSTS AND IS ALSO			
CC	REQUIRED FOR PATHOGENICITY IN HOST PLANTS.			
CC	-1- SUBCELLULAR LOCATION: SECRETED; VIA THE HRP SECRETION PATHWAY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; L78834; AAB49733.1; -			
KM	Hypersensitive response.			
FT	DOMAIN	1	224	GLY-RICH.
FT	NON_TER	356	356	
SO	SEQUENCE	356 AA;	108BA6B9D27F9DE4 CRC64;	

Query Match	39.98;	Score 828.5;	DB 1;	Length 356;
Best Local Similarity	48.28;	Pred. No. 1.4e-41;		
Matches 196;	Conservative 37;	Mismatches 115;	Indels 59;	Gaps 11.

OY 3 LNTSGAGASTWQISIGGAGGNNGIL-GTSRONAGLGGNSALGLGGNONDTVNOLAGLLT 61
 || : ||| :|| : ||| | | | | | : : : || : : :
Db 2 LNSLGGAS-LQITI-KAGCNGGLEPSSQSONGGSPQSAF---GGQRNIAEQLSIDIMT 56

QY 62 GMAHMAHMAAGGGGATMGCGLG-----GGTGGGAGGSSGAGGAGLAAHNDAGGSSGTTNTGSS 11.5

Db 57 TMMFMGSMAGCG--MSGGTGGGAGSSGAGGAGGAGG--GGTGGGAGSSGAGGAGGAGG--GG 11.1

QY 116 KCGNNTYSTTNSPLDQALGINSTSQNDSTGCTDSTSDSSDPWQQLKMFSEIQSLFGD 17.5

Db 112 LGG-----ALGAGMNAHNPNSAMNGS-----LTFSALEDLGLG 14.3

QY 176 GODGQGGSSGCGKPTBEEQNAAYKKGYTDALSGTGMNGSLGGLGNGGLGGGGGAGTGL 23.5

Db 144 GMSQGGGGLFENKQSPSSPEISATAYGQVNDALSLGNSLSQTKQ-----T 16.9

QY 236 DGSSTGGKGLONLSGPDVYQQLGNAVGTGCGKAGTQALNDIGTHRSSSTPSFVNKGDRA 29.5

Db 190 SPLQIGNNGGLGAGFNLGSLGAGFNLGSLTGSSVQKGLGLOELNNISTHNDSPRYEVDKEDRG 24.9

QY 296 MAKEIGQFMDQYPEYFVGKPOYQKPGCEVFKDDKSSMAKALSKPPDDGMPFASMEQFKAK 35.5

Db 250 MAKEIGQFMDQYPEYFVGKRAEYQKDNWQTAKEDEKSMKALSKPPDDGMPGSMQKFKKAV 30.9

QY 356 GMIRPMAGDTCNGTLQARGGSSSLGIDAMMADADINNALGKTLGA 40.2

Db 310 GMTSALRGDGTGNTNLARGGAGSLGIDAMMADDRIVNGLKTLSS 35.6

```

RESULT 3
ID HRPN_ERMCH STANDARD: PRT: 340 AA.
AC Q47276;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE HARPIN (HARPIN-ECH).
GN HRPN.
OS Erwinia chrysanthemi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium.
OX NCBI_TaxID=556;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EC16 / AC1450;
RX MEDLINE=9612740; PubMed=8589405;
RA Bauer D.W., Wei Z.M., Beer S.V., Collmer A.;
RT "Erwinia chrysanthemi harpinEch: an elicitor of the hypersensitive
RT response that contributes to soft-rot pathogenesis."
RL Mol. Plant Microbe Interact. 8:484-491(1995).
CC -1- FUNCTION: ELICITS THE HYPERSENSITIVE RESPONSE (HR) IN THE PLANT
CC UPON INFECTION. HARPIN ELICITS HR IN NON-HOSTS AND IS ALSO
CC REQUIRED FOR PATHOGENICITY IN HOST PLANTS.
CC -1- SUBCELLULAR LOCATION: SECRETED; VIA THE HRP SECRETION PATHWAY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation --
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L39897; AAC31978.1; -.
KW Hypersensitive response.
FT DOMAIN 1 203 GLY-RICH.
SEQUENCE 340 AA; 34274 MW; EAD82023731E4F4C5 CRC64;

```

Query Match	34.68;	Score 718.5;	DB 1;	Length 340;
Best Local Similarity	42.68;	Pred. No. 3.1e-35;		
Matches 172;	Conservative 41;	Mismatches 114;	Indels 77;	Gaps 10;

```

Oy 13 MOIST-GGAGGNNGLGTSRQNALGG-NSA--LGAGGNNQNDVTNQLAGLITGMMMM 677
    |||:|  ||:|:|  |||  ||  :|::|  ||  ||
Db 1 MOTTIAHIGDGLGVSGLGLGAGGLKGLNSAASSLGSSVDKLTSTIDKLTLSALTSM--- 577

```

QY 68 SMWGGGGLMGGLGGLGNGELGEGLSNALNDMLGSLNTLTGSKGNNTTSTNS 127
 DB 58 -----FGALAGGLGAS-SKGLGMSHOLGOSFEN-----GAGASNLISVPK- 98
 QY 128 PLDOLALINSTSONDSTSGTSDTSDSDPMOOLKMFSEIMOSLEF-----DQDQD 179
 DB 99 -----SGGDALS-----KMFDAALDDLGHDTVTKLTNTNSQ 130
 QY 180 TQSSSSGKOPTEGEONAYKKVYTDALSLGLMGLSGLNLGGLGAGGNGAGTGLDSS 239
 DB 131 LANSMLNASQMTQGNNAFESGVNNAISLILGNGLGOSM-----SGFSQPS 176
 QY 240 LGAGGLONTSGPYDQOLGNAVGTGIGMKAGLQALNDIGTHRHSSPTFSFYKNGDRMAKE 299
 DB 177 LGAGGLGGLSGAGAFNOLGNAIGMGVGNALSLNSVTHVDGNHNEFDKEDRGMAKE 236
 QY 300 IGFMDQYPEVFGKPOYOKGPGQEVKTDGSKMAKALSKPDDGWTTPASMEQFNKAKMIK 359
 DB 237 IGFMDQYPEIFGKPEYOKGWSGSPKTDGSKMAKALSKPDDGWTTPASMEQFNKAKMIK 296
 QY 360 RPAAGTGNGNTLQKAGAGSSSLGIDAMAGDAINNALGKLGAA 403
 DB 297 SAVAGDTGNTNMLRGAGASLGDIAAVVDGDKIANMSLGKLANA 340

RESULT 4

YAG8_MYCTU STANDARD; PRT; 463 AA.

AC 053416;
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DE HYPOTHETICAL PE-PGRS FAMILY PROTEIN RVL1068C.
 GN RVL1068C OR MTVO17.21C.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 NCBI_TaxId=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
 RA Taylor K., Whitehead S., Barrell B.G.;
 RA Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RL Nature 393:537-544(1998).
 CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGRS
 SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: AL021897; CA117184.1; -
 DR Tuberculist: Rv1068C; -
 DR InterPro: IPR000084; -
 DR InterPro: IPR002952; -
 DR Pfam: PF00934; PE; 1.
 DR PRINTS: PRO1228; EGSSEHL.
 KM Hypothetical protein.
 SQ SEQUENCE 463 AA; 39305 MW; CE5696A7E9593952 CRC64;

Query Match 11.98; Score 248; DB 1; Length 463;
 Best Local Similarity 31.88; Pred. No. 6,7e-08;
 Matches 93; Conservative 17; Mismatches 112; Indels 70; Gaps 13;

QY 8 LGASTMOISIGAGGNGGLGTSTRONAGLSNALSILGSGGNQNDYVQLGLTLGMMMM 67
 DB 203 IGAPVAGAGAGAGAGTAGLGG-----KCGAGAGAGAGAGAGRGD--GSGAGWLSGNGGDA 256
 QY 68 SMWGGGGLMGGLGGLGNGELGEGLSNALNDMLGSLNTLTGSKGNNTTSTNS 127
 DB 257 GTGGGGGNAGGNGGAGWLSGNGGTGGG-----GSTAGA-GGGGNG-----NS 301
 QY 128 PLDOLALINSTSONDSTSGTSDTSDSDPMOOLKMFSEIMOSLEFDDGQDGTGSS--- 184
 DB 302 GIDPG---NGOGADTGNAGCGHGSA-----AKLFDGAGAGAGGNGST 344
 QY 185 -----SGKQPTGEONAYKKVYTDALSLGLMGLSGLL-----GNGGLGGGQ--G 229
 DB 345 GGTGGGGGFGGCTGCGNGGNGHAGAGSGGTAGLGSAGSGCTGGDGGNGGLGSGCAKG 404
 QY 230 NAGTGLDSSLGKGLQNTSGPYDQOLGNAVGTGIGMKAG---TQALNDIG 278
 DB 405 NGNGGD-----GKG-----GDAQLGNGGNGGNGKGTGLMPGINTG 445

RESULT 5

Y034_MYCTU STANDARD; PRT; 778 AA.

AC P71933;
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DE HYPOTHETICAL PE-PGRS FAMILY PROTEIN RV2634C.
 GN RV2634C OR MTCY441.04C.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 NCBI_TaxId=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
 RA Taylor K., Whitehead S., Barrell B.G.;
 RA Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RL Nature 393:537-544(1998).
 CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGRS
 SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: Z80225; CAB02341.1; -
 DR HSSP: P04002; IATF.
 DR Tuberculist: Rv2634C; -
 DR InterPro: IPR000084; -
 DR InterPro: IPR001899; -
 DR Pfam: PF00934; PE; 1.
 KM Hypothetical protein.

SQ SEQUENCE 778 AA: 63131 MW: DAB20FE58E4999E7 CRC64:

Query Match 11.6%; Score 242; DB 1; Length 778;
Best Local Similarity 30.1%; Pred. No. 2.6e-07;
Matches 91; Conservative 19; Mismatches 110; Indels 82; Gaps 14;

```

OY 7 GIGASTMOISIGAG--GNNGLGTSRONAGIGNSALGLAG-----GNQNDTVNOLAGLL 60
DB 353 GNGGNANMFSGAGGCGGTGAGTNGVNPISIANPNTGANCNDNSNGNQ----- 404
OY 61 TGMNMMNMMGCGGLMGGLGGLG--NGLGSSGGLGEGLSNALDMLGSLNTLGSKG 118
DB 405 -----GNG--GGPPAGVGAGVGAGGGLGEGSLDG--MDGTGG-----KGG 443
OY 119 NNTTSTNSPLDQALGINSTSONDSTSGTSDTSDSPMQLLMKFSEIMQSLFGDGD 178
DB 444 AGGTAGTGGAGGAGGAGGAGIGTDSAGV-----ATGEGGD 481
OY 179 GTGSSSGGKOPTGE--ONAYKKVTDAL---SGLMNG---LSQLLNGGLG----- 225
DB 482 GATGVDGVDGAGGAGGAGGAGHTGVDAGFGGCGIGGDNALGAAGNGGTGAGGNGG 541
OY 226 -----GGGNAGTGLDSSLGKGLNLSGPV---DYQLGNANGTGTGMAAGTQALND 276
DB 542 RCGMLIGNGAGGAGGTGT--GGGAGAGFAGVGAGGAGGGLTMDAGTGTGGLGGLG 600
OY 277 IG 278
DB 601 VG 602

RESULT 6
Y278_MYCTU STANDARD: PRT: 957 AA.
AC P56877:
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE HYPOTHELTICAL PE-PGRS FAMILY PROTEIN RV0278C PRECURSOR.
GN RV0278C OR MTV035.06C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV:
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagers K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter K., Whitehead S., Barrell B.G.;
RA Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
-1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGRS
SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch)
CC
CC EMBL: AL021930; CAA17353.1;
CC
CC TubercuList: RV0278C;

```

DR InterPro: IPR000084; -
DR Pfam: PF00934; PE; 1.
RN Hypothetical protein; Repeat; Signal.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 957 HYPOTHELTICAL PE-PGRS FAMILY PROTEIN
FT RV0278C.
SQ SEQUENCE 957 AA: 81905 MW: 71EBAD417FBA47C CRC64:

Query Match 11.6%; Score 241.5; DB 1; Length 957;
Best Local Similarity 25.8%; Pred. No. 3.4e-07;
Matches 117; Conservative 42; Mismatches 149; Indels 145; Gaps 21;

```

OY 2 SLNTSLGASTMOISIGAGNNGLGTSRONAGLAG-----NSALGLGQNDTVNOLA 57
DB 362 TLTAGIGGA-----GGAGNAGLLFGSGSGAGAGGREFADGGGGRGN-----A 407
OY 58 GLITGMMNMMNMMGCGGLMG--GGLGGLGNGLGSSGL-----GGLSNAL--N 103
DB 408 G-----TFVSGSGAGGNGGVGGFAGGAGTGTGGLGNGNGNGGASAVTGN 458
OY 104 DMLGSLNTLGSKGNNTTSTNSPLDQALGINSTSONDSTSGTSDTSDSPMQLLK 163
DB 459 GGIGGTGVLGN--GGNGSGGIGAGKAGVGVGGLLGLDGFNAPASTSPHTLQNNVLN 517
OY 164 MFSEIMQS-----LFGDGDGTGGS---SSGKOPTGEQNAAYKKVTDALSG--LMNG 213
DB 518 VYNEPQTILTGRLPILGNANGTPTGTAGDAGAGWLEGNANGTPTGTGAAGAGWLFNG 577
OY 214 LSQLLNGGLG-----GGGNAGTGLDSSLGKGLNLSGPV 254
DB 578 -----GNGHGATNTAATATGAGAGAGGILFCTGNGGTG--GIATGAGGIGAGAGAGV 630
OY 255 QQLGNAYGT-----IGKKAGIQALNDIGTRHSSTRSFVKKGRRAKELGQPMQYR 308
DB 631 SLILSGSGGTGNGKNSIGV--AGIGGAGRG-----GDGL----- 664
OY 309 EYFGKPYQYKPGQEVKTTDKSMARALSKPDDGMTPASMEQFNFKAKIKRPMAGDTGN 368
DB 665 -LFG-----AAGTGG-----HGAAGVPAVGAGAGGN 690
OY 369 GMLQARGAGSSSLGIDAMMAGDAINNALGKLG 401
DB 691 GGLFANGAGAGAGFMA--AGNGNGGLFGTGG 722

```

```

RESULT 7
Y140_MYCTU STANDARD: PRT: 515 AA.
AC Q50594:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE HYPOTHELTICAL PE-PGRS FAMILY PROTEIN RV1840C.
GN RV1840C OR MTY1A11.04 OR MTY3359.33.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV:
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagers K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter K., Seeger K., Skelton S., Squares S., Sgares R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."

```

RL Nature 393:537-544(1998).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGRS
 CC SOBEFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z83859; CAB06114.1; -
 DR Tuberculin; RV1840C; -
 DR InterPro: IPR000084; PE; 1.
 DR Pfam: PF00934; PE; 1.
 KM Hypothetical protein; Transmembrane.
 FT TRANSMEM 1 21 POTENTIAL.
 FT TRANSMEM 165 185 POTENTIAL.
 FT TRANSMEM 199 219 POTENTIAL.
 FT SEQUENCE 515 AA; 43916 MW; 2EDC8B6C6D28B7E3 CRC64;
 SQ
 Query Match 11.5%; Score 238.5; DB 1; Length 515;
 Best Local Similarity 27.3%; Pred. No. 2.7e-07;
 Matches 109; Conservative 23; Mismatches 128; Indels 139; Gaps 15;
 QY 18 GGAGGNNGLTSTSRONAGLSALGLGQNDVNOVLGLTNGMMAMSMGSGGLMG 77
 DB 148 GGAGGADGLGNG-GNGGIGPGATGLAG-----AGVGGLLKGDDGNGAGGLG 197
 QY 78 GGLGGGGLNGLSGGGGLGGLSNALNDMLGSLNTLSKGGNNTTSTNSPL--DQALGI 135
 DB 198 TGPVGATG-GIGGPGAGVAVGLFGHGAGCAGAGGLKAGAGAGTGTGGLTYGNGNGG 256
 QY 136 NSTSONDSTGTSTSSDPMDQLKMFSEIMQSLFGDDODDQSGSSSGKOTBEBO 195
 DB 257 NVPSGADGAGGDA-----RLIGNGDG---GSVCAATPG--- 289
 QY 196 NAYKKVTDALSGLMGSLGLGNGGLG--GCGNAGTGLDSSSLGSKGLQMLSPV 252
 DB 290 -----IGNG-----GNGGNGMLYGDGSGGSLTGFSDDGTG----- 322
 QY 253 DYQOLGNAVGTGIGKACIGIALNDIGTRHHSSTRSFVKKGDRAAKETIGQFMDOYPEVFG 312
 DB 323 -----GNMGMGDDGNGGFFDGGGDDGTGTLIGNGD----- 358
 QY 313 KPOYQKPGQEVKTDKMAKALSKRPDDGMPASMEQFNKAKMIKPMAGDTGN--- 368
 DB 359 -----GNSVQTD-----GFL-KHGHDGDNNAVGL 382
 QY 369 -GNLQARGAGSSSLGIDAMAGDAIIN---MALKLG 402
 DB 383 ICGNAGAGAGAGTGVAPAGSGSGGNGGALLVNGCA 421
 RESULT 8
 Y208 MYCTU STANDARD; PRT: 1901 AA.
 AC OS3553;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DE 30-MAY-2000 (Rel. 39, Last annotation update)
 DE HYPOTHEICAL PE-PGRS FAMILY PROTEIN RV3508 PRECURSOR.
 GN RV3508 OR MTW023.15.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-H37RV;

RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hounsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
 RA Taylor K., Whitehead S., Barrall B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RA complete genome sequence.";
 RL Nature 393:537-544(1998).
 CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGRS
 CC SOBEFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AL022022; CAAL1745.1; -
 DR HSSP: P19972; IKVD;
 DR Tuberculin; RV3508; -
 DR InterPro: IPR000084; -
 DR Pfam: PF00934; PE; 1.
 KM Hypothetical protein; Repeat; Signal.
 FT SIGNAL 1 30 POTENTIAL.
 FT CHAIN 31 1901 HYPOTHEICAL PE-PGRS FAMILY PROTEIN
 FT SEQUENCE 1901 AA; 147627 MW; C7B1923D5D0146CD CRC64;
 SQ
 Query Match 11.4%; Score 238; DB 1; Length 1901;
 Best Local Similarity 27.6%; Pred. No. 1.1e-06;
 Matches 124; Conservative 29; Mismatches 180; Indels 116; Gaps 23;
 QY 9 GASTMOISI-GGAGGNNGLTSTSRONAGLSALGLGQNDVNOVLGLTNGMMAM 67
 DB 1133 GASTTISINANGAGAGNGGIGG---KGGAGAGTGLTVGSSGTTGGDGGAG----- 1179
 QY 68 SMNGGGLMGGLG--GGLGN-GLGSSGLG--EGLSNALNDMLGSLNTLSKGGNNTTS 123
 DB 1180 --SGGGGPGFAGAGAGAGGNGGNGGVDGGEASGLGLISCFDGGQ---GGGGAGGSA 1233
 QY 124 TTNSPLDQALGINSTSQNDSTSGDSTSDSPMDQLKMFSEIMQSLFG--DGODGTQ 181
 DB 1234 -----GAGGINGAG---GAGGTGGAGCGCAP-----ATLIGPDDGGDGO 1270
 QY 182 GSSSGKQPTGEONAYKKGVY---DALSGLMGSLGLGNGGLG--GGGNGNAGTGL 235
 DB 1271 GGIGG-----DGNMGFGAGVPGDGGDGNAGFAGVP---GDGIGTGAGAGAGAGA 1322
 QY 236 DG-----SSLGKGLNLGSPVDYQO-----LGNA 260
 DB 1323 DCDPSIDGGGAGAGGCHGGCGGKGLNSTGLASAAGSDGGNGAGAGGNGGDDGDTIGGS 1382
 QY 261 VGT-GTGMKAGIOAL-NDIGTRHHSSTRSFVKKGDRAAKETIGQFMDOYPEVFGKPOYK 318
 DB 1383 GGTGGTGDRAGVGGGLANTGTCGNMGTGAGAGRGDGGGAGSALSDGNGFRAGGGGGG 1442
 QY 319 GPQOEKVTDDKMAKALSKRPDDGMPASMEQFNKAKMIKPMAGDTGN---GDTGNGMLQAR 374
 DB 1443 GVGNGNAGAGINGAGTG---GTGAGAGDGGNGTGTVAASBAGAGGGGDDGGGGGT--- 1494
 QY 375 GAGGSSLGIDAMAGDAIINMALKLGAA 403
 DB 1495 GGAGGNAAGFAGVPGD---GIGGTGGA 1519
 RESULT 9

WA22_MYCTU STANDARD: PRT: 914 AA.
 ID WA22_MYCTU
 AC 006794;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE WAG22 ANTIGEN PRECURSOR.
 GN WAG22 OR RV1759C OR MYC28.25C.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 NCBI_TaxID=1773;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed-9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
 RA Taylor K., Whitehead S., Barrell B.G.;
 RT Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RL Nature 393:537-544(1998).
 CC -1 SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGSS
 SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: 295890; CAB09322.1; -
 DR HSSP: P41140; 2SFA.
 DR Tuberculist: RV1759C; -
 DR InterPro: IPR000084; -
 DR Pfam: PF00934; PE: 1.
 DR Antigen; Repeat; Signal.
 FT SIGNAL 1 30 POTENTIAL.
 FT CHAIN 1 914 WAG22 ANTIGEN.
 FT SEQUENCE 914 AA; 74354 MW; F6953C3DBBE6AC8 CRC64;
 SO
 Query Match 11.3%; Score 235.5; DB 1; Length 914;
 Best Local Similarity 25.8%; Pred. No. 7.2e-07;
 Matches 120; Conservative 38; Mismatches 173; Indels 135; Gaps 20;
 QY 7 GICASTMQLSICGAGCNG-----LIGTSRONACLGNSAL-----GLCGNQNDTVNQ 55
 DB 186 GAGGASAMLCGAGAGAGAGATSLVGICTGCTGAGNACMLAGAGAGAGGCSFSTAGC 245
 QY 56 LAGL-LTGMMMMNMVMGGGGLMG-----GGUGGGLG-NGUGSGGGLG-----GLSNA 101
 DB 246 AGGAGAGAGLFTTGGVGGAGGCGHTGTGAGAGAGAGLFGAGGAGGAGGFGDHTGLGTGA 305
 QY 102 LNDMLGSLNTLIGSKGNNNTSTTNSPLDQALGINSTQNDSDSTGTDSTSDSDPMQOL 161
 DB 306 GGGGGGGGLFGAGGAGGAGGSLTTGA-AGNGNAGTLLGAAAGAGCGTGGAGC----- 359
 QY 162 LKMFSEIMSLFEDGDDGTGCS-----SSGKOPTBEEDQNNYKKGVTDAISGLMGNG 213
 DB 360 -----TVFGGKAGGAGAGAGNMLFGSGGGGCTGG-----FGFAAGCGGAGGGS 404
 QY 214 LSQLNGGGLGAGGCG-----GNAGTGLDSSLSLGGKGLQNLNSGP 251
 DB 405 AGMLSSGGSGGAGGAGPACTAAGAGAGAGAPGLIGNGNGNGGSGGTGGVAGAG- 463

QY 252 VDYQOLGNAVGTGIGMKAGIOA-----LNDIGTRHSSTRSFVAKGDRAMA 297
 DB 464 -----GNAYLIGNGGEGGIGALAGKSGFCGGLLGGADYNAPESTSPWNH-----LQ 512
 QY 298 KEIGQPMDDIPEYF-GKPOYQKPGQEVKTDDKSWAKALSPDDCKMTPAEMQFNKAG 356
 DB 513 QDILSFTEPTALTGSPPLGNG-----DSQTPGTGDD-GGAGC 550
 QY 357 MIRKPMADTGNGLQARGGSSSLGIDAMAGDATNNMALGRTGA 402
 DB 551 W-----LFGNGNGAGAGAGTNGSAGC-----ACGA-GGILFTGGA 586
 RESULT 10
 ID YP91_MYCTU STANDARD: PRT: 543 AA.
 AC YP91_MYCTU
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-MAY-2000 (Rel. 39, Last annotation update)
 DE HYPOTHETICAL PE-PGSS FAMILY PROTEIN RV2591.
 GN RV2591 OR MYC27.10C.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 NCBI_TaxID=1773;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed-9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
 RA Taylor K., Whitehead S., Barrell B.G.;
 RT Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RL Nature 393:537-544(1998).
 CC -1 SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGSS
 SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: 277724; CAB01283.1; -
 DR Tuberculist: RV2591; -
 DR InterPro: IPR000084; -
 DR Pfam: PF00934; PE: 1.
 DR Hypothetical protein.
 FT SEQUENCE 543 AA; 46287 MW; 59730339E5D2DE59 CRC64;
 SO
 Query Match 11.3%; Score 234.5; DB 1; Length 543;
 Best Local Similarity 28.0%; Pred. No. 4.8e-07;
 Matches 111; Conservative 31; Mismatches 121; Indels 133; Gaps 21;
 QY 18 GAGAGNNGLIGTSRONAGLGCNSALGIGGQNDTVNQLAGLGTGMMMMNMVMGGC----- 72
 DB 213 GAGAGDAPLLIGWG-GNGGPGGFAFGNGAGAGN-----GASGSLFGVGGAGGAGGSS 264
 QY 73 -----GGLMG-GGLGGGLGNGLGGSGGLGGLSN-----ALNDMLGSLNTLIGSKGNNNT 122
 DB 265 EDVGCTGAGAGAGAGLGLGAGGAGAG-GTSSNNNGGAGGAGTAGGGLFSLGAGGAGGGA 323
 QY 123 STTNSPLDQALGINSTQNDSDSTGTDSDSDPMQOLLKMFSEIMQSLFGDQDGTG 182


```

Db      324 GT-----AIGSNVA-----GDGAGAGDSS-----ALIGTAQGGSGG   353
OY      183 SSSGKOPTBEEONAKKYCTDALSGLMGNLSOLLINGGLGGGQ---GCNATGIDGSS   239
Db      354 --LGG--FEESTGGDDGLGAGAVILICTGVG--GGGIGGGSGNGTGGAGAGTGAT   404
OY      240 L-----GGKQLNLISGPVDYQOLGNANVTGICKKAGI-QALNDIGTHRHSSSTSPVNGK   292
Db      405 LIYGAGAGGGGIGCPA-----VNWNGNVG-GLGGCGGGGALLIGLAG-----GAG   449
OY      293 DRAMAKEIGQMDPYPEVFGRPOYOKGGGEFKTDKSWAKALKSPDDGATPAMEOFN   352
Db      450 GAGGATVVG-----LGNGSDG-----G   467
OY      353 KAKGMIRPMAGD---TNGNMLQARGAGGSSLGIDA   385
Db      468 DGGGLFSIGVGDDGNMANGAMPANGGNGAGVYA   503

RESULT  11
GRP_ARATH STANDARD; PRT; 338 AA.
AC      P27483;
DT      01-AUG-1992 (Rel. 23, Created)
DT      01-AUG-1992 (Rel. 23, Last sequence update)
DT      01-AUG-1992 (Rel. 23, Last annotation update)
DE      GLYCINE-RICH CELL WALL STRUCTURAL PROTEIN PRECURSOR.
OS      Arabidopsis thaliana (Mouse-ear cress).
OC      Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC      Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC      Brassicales; Brassicaceae; Arabidopsis.
OX      NCBI_TaxID=3702;
RN      [1]
RP      SEQUENCE FROM N.B.A.
RC      STRAIN-CV. COLUMBIA:
RX      MEDLINE-92003708; PubMed-1912511;
RT      Outgley F., Villiot M.L., Maché R.;
RA      "Nucleotide sequence and expression of a novel glycine-rich protein
    gene from Arabidopsis thaliana."
RL      Plant Mol. Biol. 17:949-952(1991).
CC      -1- FUNCTION: RESPONSIBLE FOR PLASTICITY OF THE CELL WALL (POTENTIAL).
CC      -1- SUBCELLULAR LOCATION: CELL WALL (POTENTIAL).
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@lsb-sib.ch).
CC      -----
DR      EMBL; X58338; CAA41249.1; -.
DR      PIR; S17732; KMMU.
DR      HSSP; P30129; 4DPV.
FM      Cell wall; Structural protein; Repeat; Signal.
FT      SIGNAL 1 20
FT      CHAIN 21 338 GLYCINE-RICH CELL WALL STRUCTURAL
FT      PROTEIN.
DOMAIN 21 338 GLY-RICH.
SEQUENCE 338 AA; 23891 MW; 046A6E8C1A4E89EB CRC64;

```

Query Match	11.2%	Score 233.5	DB 1:	length 338
Best Local Similarity	29.9%	Pred. No. 3	3e-07	
Matches	80	Conservative	14	Mismatches 77; Indels 97; Gaps 10
OY	7	GIGASTMQISTIGAGGANNGLGTSKRNA--GGGASNALGAGGAGNNDYNNOLAGLLTGM	64	
DB	59	GG-----GGAGGGGGGGAGGAGGAGGAGGAGGGGGGGH-----GGGGG--	100	
OY	65	KMSHMGCGGGLGCGGAGGGGAGNGLG--SGGIGEGCLSNALNDMLGGLSLTLTASKRGGNNTT	122	

```

Db      101  -----GAGGAGAGGGLGGGHHGGGJGGGAGGGGGGGGLGGGAGGGAGGGGGGGLGGGH-- 152
Qy      123  STTNSPLDQALGINSTSNDSTSGTSTSSSDPMQQLKMFSEIMQSLFGCGQDGTG 162
Db      153  -----GGGIGGGAG 161
Qy      183  SSSGCKPTEGEQNAVKKGYTDALSGLMGNCLSQLLNGNGGLGGGGGAGNAGTGLDSSILGG 242
Db      162  GGAGG-----GLGGHGGCGTGGGAGGSGGGGL-----GGGIGGAGAGGAGGG--GAGGG 209
Qy      243  KGLQNTSGPVDYQQLGNAVGTGIGKAG 270
Db      210  GGLGGGHH-----GGRGGGAGGGLGGGAG 233

```

```

12 RESULT
13 YK98_MYCTU
14 ID YK98_MYCTU STANDARD: PRT: 434 AA.
15 AC Q10707;
16 DT 01-OCT-1996 (Rel. 34, Created)
17 DT 15-JUL-1999 (Rel. 38, Last sequence update)
18 DT 30-MAY-2000 (Rel. 39, Last annotation update)
19 DE HYPOTHEETICAL 36.5 KDA GLYCINE-RICH PROTEIN RV2098C.
20 GN RV2098C OR MYCY49.38C.
21 OS Mycobacterium tuberculosis.
22 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
23 OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
24 OX NCBI_Taxid=1773;
25 RN [1]
26 RP SEQUENCE FROM N.A.
27 RC STRAIN=H37RV;
28 RX MEDLINE=98295987; PubMed=9634230;
29 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
30 Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
31 Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
32 Davies R., Dellin K., Feldwell T., Gentles S., Hamlin N., Holroyd S.,
33 Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
34 Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
35 Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
36 Taylor K., Whitehead S., Barrett B.G.;
37 RA *Deciphering the biology of Mycobacterium tuberculosis from the
38 complete genome sequence."
39 RL Nature 393:537-544(1998).
40 CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGRS
41 SUBFAMILY.
42 CC -1- CAUTION: THIS PROTEIN IS FRAMESHIFTED NEAR THE N-TERMINUS.
43 CC THE SEQUENCE HAS BEEN CHECKED BY AUTHORS IN REF.1 AND THEY REPORT
44 THAT NO ERRORS HAVE BEEN FOUND.
45 CC -----
46 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
47 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
48 CC the European Bioinformatics Institute. There are no restrictions on its
49 CC use by non-profit institutions as long as its content is in no way
50 CC modified and this statement is not removed. Usage by and for commercial
51 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
52 CC or send an email to license@isb-sib.ch).
53 CC -----
54 DR EMBL: Z73966; CAA98228.1; -
55 DR HSPF: P19972; 1KVD.
56 DR TuberculinSt: RV2098C; -
57 DR InterPro: IPR000084; -
58 DR InterPro: IPR002952; -
59 DR Pfam: PF00934; PE. 1.
60 DR PRINTS: PR01228; EGGSHLL.
61 DR Hypothetical protein.
62 SQ SEQUENCE 434 AA; 36530 MW; B736E822CD1E0E54 CRC64;

```

Query Match	11.2%;	Score 233.5;	DB 1,	Length 434;
Best Local Similarity	28.8%;	Pred. No. 4.3e-07;		
Matches 105;	Conservative 22;	Mismatches 150;	Indels 87;	Gaps 16
QY	9	GASTMGIISIGAGGNNCLCTSRQNMGLGGSNALSIGGGGNQNPVNMAGLL-----	60	

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: 273902; CAA8089.1; -
 DR HSSP: P19972; IKVD.
 DR Tuberculist; RV1325C; -
 DR Interpro; IPR000084; -
 DR Pfam; PF00934; PE: 1.
 KM Hypothetical protein: Repeat; Signal.
 FT SIGNAL 1 30
 FT CHAIN 31 603
 FT DOMAIN 114 603
 FT SEQUENCE 603 AA; 49575 MW; 4F9BCB82B07AE964 CRC64;

Query Match 11.08; Score 229.5; DB 1; Length 603;
 Best Local Similarity 26.38; Pred. No. 1e-06;
 Matches 114; Conservative 32; Mismatches 132; Indels 155; Gaps 24;

QY 1 MSIN-----TSGASTMOTISIGAGGNN--GLGTSRONAGLGNL-----GLGGGN 48
 DB 262 LSVNCGTGTGGTGGGGLTSGAGAGGFGVSGSAGNGTGGDGIPTNGTGTGTG 321
 QY 49 QNDVTNOLAGLGTMMAMSMAGGG--LNG-GGLG--GLG-----NLGSGSGL 94
 DB 322 GTGTGNLGVGEGS-----AGGAGNAGILFAGAGTGTGTGTGTGTGTGTGTGTG 374
 QY 95 GEGSLNLMNMGSLTLTSGKGNNTTSTNSPLDALGINSTSNDSTSTSTSDS 154
 DB 375 G-GIGCA-----GALRPGAGAGGTG-----GFGASSADQMAGIGSGSGGA 416
 QY 155 SDPMQQLMKFSEIMOSLFGDQDGTGSSSGKQPTGEBQNAKKYTDALSLMGGL 214
 DB 417 A-----KLIGGGAG--GTG-----DSVRGAAGG- 440
 QY 215 SOLGNGGLG-----GGGQGNAGTGLDSSILGKGLNLGSPVDYQQLGNV--TGISM 268
 DB 441 ----GTGTGGLIDGAGAGGTGIERGSVAGAGAG-----GNAAGLSGAGGA 486
 QY 269 AGIOALNDIGTHRHSTRSEYVNGKDRAMAKETIGFMDQYPEVFGKPOYQKPGQEVKTDD 328
 DB 487 GGAGGFG-----TAGDGGAGNAGLL-----NGDG-----G 513
 QY 329 KSNMALKSPDDDDMTASMPQFNKAKMIRPMAAGTGNGLDARGAGSSSLGIDAMMA 388
 DB 514 AGGAGGGLGAGIDG-----NGKRGK--KAGVANGNGDG-----GAGGASVAVANGVG 558
 QY 389 GDALNNMALKIG 401
 DB 559 GSGGNATLIGNG 571

RESULT 15
 GRP1_PETHY STANDARD; PRT: 384 AA.
 AC P09789;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 01-AUG-1992 (Rel. 23, Last annotation update)
 DE GLYCINE-RICH CELL WALL STRUCTURAL PROTEIN 1 PRECURSOR.
 GN GRP-1.
 OS Petunia hybrida (Petunia).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;
 OC Solanales; Solanaceae; Petunia.
 OX NCBI_TaxID=4102;
 RN [1]

RP SEQUENCE FROM N.A.
 RA Condit C.M., Meagher R.B.;
 RT "A gene encoding a novel glycine-rich structural protein of petunia.";
 RL Nature 333:178-181(1986).
 CC -1- FUNCTION: RESPONSIBLE FOR PLASTICITY OF THE CELL WALL (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: CELL WALL (POTENTIAL).
 CC -1- MISCELLANEOUS: THIS PROTEIN CONTAINS 678 GLYCINE RESIDUES.
 CC -1- MISCELLANEOUS: 90% OF THE MATURE PROTEIN RESIDUES ARE CAPABLE OF
 CC FORMING A BETA-PLEATED SHEET COMPOSED OF 8 ANTI-PARALLEL STRANDS.
 CC -1- MISCELLANEOUS: THE GLYCINE-RICH REGION IS COMPRISED OF TWO RELATED
 CC FAMILIES OF REPEATS, F1 AND F2, EACH REPEAT CONTAINING ABOUT 40
 CC AA.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: X04335; CAA27866.1; -
 DR PIR; A26099; A26099.
 DR HSSP; P30129; 4DPV.
 KW Cell wall; Structural protein: Repeat; Signal.
 FT SIGNAL 1 27
 FT CHAIN 28 384
 FT DOMAIN 41 384
 FT SEQUENCE 384 AA; 28777 MW; C8541C549417D18C CRC64;

Query Match 10.38; Score 214.5; DB 1; Length 384;
 Best Local Similarity 28.18; Pred. No. 4.8e-06;
 Matches 77; Conservative 17; Mismatches 111; Indels 69; Gaps 9;

QY 7 GLASTMOTISIGAGGNNGLGTSRONAGLGNLGLGAGGNNQNDVTNOLGLTMM 66
 DB 152 GFGAG-----GVGGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 201
 QY 67 MSNMGGLMGGGLGGLGNGLGGSGGLGEGLSNLMNMGSLTLTSGKGNNTTSTTN 126
 DB 202 ----GAGGGLGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 243
 QY 127 SPLDALGINSTSONDSTSTSTSDSDPMQQLMKFSEIMOSLFGDQD-----GTQ 181
 DB 244 ----GVGGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 280
 QY 182 GSSSGKQPTGEBQNAKKYTDALSLMGNGLSQLGNGLGGGAGGAGGAGGAGG 239
 DB 281 GGVGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 335
 QY 240 LGKGLNLSGPDYQQLGNVAGTIGMKAGIOA 273
 DB 336 GGGGIGGGHGG-----GFGVGVGIGIGVGA 363

Search completed: October 25, 2001, 11:47:30
 Job time: 191 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 25, 2001, 11:43:59 ; Search time 24.18 Seconds
(without alignments)
2205.084 Million cell updates/sec

Title: US-09-412-100-23
Perfect score: 2079
Sequence: 1 MSMTSGAGASTMQISIGTA.....DAMAGDAINNALGKLGAA 403

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues
Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP:REMBL.16:*
2: sp.archaea:*
3: sp.bacteria:*
4: sp.fungi:*
5: sp.human:*
6: sp.invertebrate:*
7: sp.mammal:*
8: sp.mhc:*
9: sp.organelle:*
10: sp.phage:*
11: sp.plant:*
12: sp.podent:*
13: sp.sp.vertebrate:*
14: sp.virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1181.5	56.8	382	2	09FCY8
2	1002.5	48.2	365	2	09KH45
3	736.5	35.4	344	2	09EXPO
4	712	34.2	142	2	09LAW6
5	264.5	12.7	1489	2	053559
6	255	12.3	591	2	007224
7	249	12.0	731	2	050415
8	248	11.9	1381	2	053552
9	247	11.9	588	2	050336
10	246.5	11.9	1079	2	053557
11	246	11.8	1538	2	053395
12	245.5	11.8	1660	2	053215
13	245	11.8	923	2	053890
14	238.5	11.5	882	2	053845
15	236.5	11.4	714	2	053556
16	236.5	11.4	853	2	053439
17	235	11.3	439	2	006287
18	233.5	11.2	349	10	09LSP2
19	233.5	11.2	615	2	005806

20	232	11.2	837	2	053684	053684 mycobacteri
21	229.5	11.0	1329	2	006810	006810 mycobacteri
22	228.5	11.0	667	2	053415	053415 mycobacteri
23	228.5	11.0	904	5	076271	076271 mytilus edu
24	228	11.0	783	2	053809	053809 mycobacteri
25	227	10.9	496	2	053631	053631 mycobacteri
26	226.5	10.9	491	2	006818	006818 mycobacteri
27	226.5	10.9	576	2	P71664	P71664 mycobacteri
28	226	10.9	694	2	053212	053212 mycobacteri
29	225.5	10.8	1011	2	P94986	P94986 mycobacteri
30	225	10.8	749	2	053844	053844 mycobacteri
31	224	10.8	1306	2	053775	053775 mycobacteri
32	223	10.7	562	2	050458	050458 mycobacteri
33	222	10.7	272	5	09VJ88	09VJ88 drosophila
34	222	10.7	767	2	053435	053435 mycobacteri
35	220.5	10.6	396	10	063450	063450 arabidopsis
36	220.5	10.6	484	2	053394	053394 mycobacteri
37	220	10.6	532	2	006215	006215 mycobacteri
38	219.5	10.6	741	2	006808	006808 mycobacteri
39	218.5	10.5	390	10	09M3Y2	09M3Y2 tritium ae
40	218.5	10.5	525	2	033284	033284 mycobacteri
41	218.5	10.5	639	2	053952	053952 mycobacteri
42	217.5	10.5	210	10	041187	041187 arabidopsis
43	217.5	10.5	220	10	039367	039367 brassica ol
44	213.5	10.3	584	2	053575	053575 mycobacteri
45	212.5	10.2	370	2	053159	053159 mycobacteri

ALIGNMENTS

RESULT 1
ID 09FCY8 PRELIMINARY; PRT; 382 AA.
AC 09FCY8;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE HRPN.
GN HRPN.
OS Erwinia stewartii.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
CC Pantoea.
NCBI_TaxID=66271;
RX [1]
RN SEQUENCE FROM N.A.
RC STRAIN=SS104;
RA Frederick R.D., Ahmad M., Majerczak D.R., Arroyo-Rodriguez A.S.,
Manulis S., Coplin D.L.; of the Pantoea stewartii subsp. stewartii hrp
RT gene cluster and sequence analysis of the hrpA, hrpC, hrpN and wtsE
RT operons.";
RL Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF282857; AAC01466.1; -
SQ SEQUENCE 382 AA; 37878 MW; 8BA8C723F2DA38BB CRC64;

Query Match 56.8%; Score 1181.5; DB 2; Length 382;
Best Local Similarity 62.1%; Pred. No. 4.7e-77;
Matches 251; Conservative 44; Mismatches 86; Indels 23; Gaps 11;
QY 1 MSMTSGAGASTMQISIGAGGNNGLTGRONAGLGNALGLG-GGNQNDTVQNLACL 59
DB 1 MSMTNSPLQTSALQVTL---GGNNGLMGTDLTDTGIGLISQPLGEGKGNBSIDLAA 57
QY 60 LFGMMNMNSMGGGGGLGGLGNGLGGSGGEGLSNLTNMLGSLNTLTGSKGN 119
DB 58 LFGMMNMNSMGGGGGL-SSLASGTGGMNSPFGGSGSAPGNTLSGTG-----GSPG- 109
QY 120 NTTSSTNSPLDQALGINSTQNDSTSGTDSDDSPMQOILKMFSEIMQSLFDDGQDG 179
DB 110 -TTGAGSS-----LGLDPTQTGDDDSLGSAGQTSGMS-PMEQLMKTFADITQSLFDD-QDG 161

```

OY 180 TCGSSGGKOPTGEBQNAKKGVTDALSGIANGSLQILGNGSLGCGGCGAGTGLDSS 239
    ID 09K45 PRELIMINARY; PRT; 365 AA.
DB 162 ASGNA-GROPQSDQBNAYKKGVTDALFAFMGGGLSQVANGSEGLDGG-GLG-CGNG 218
OY 240 LGGKGLQNLISGVVDYQOQGNVGTGIGKAGIGIOLANDITGTHRSSTRFVNGKDRAMAE 299
    ID 09K45 PRELIMINARY; PRT; 365 AA.
DB 219 LGGKGLQNLISGVVDYQOQGNVGTGIGKAGIGIOLANDITGTHRSSTRFVNGKDRAMAE 278
OY 300 IGOFPMDQYPEVFGKPOYOKPGCGEVTGTDKSWAKALSKPDDGMPASMEQFNKAKGMK 359
    ID 09K45 PRELIMINARY; PRT; 365 AA.
DB 279 VGOFPMDQYPEVFGKPOYOKPGCGEVTGTDKSWAKALSKPDDGMPASMEQFNKAKGMK 338
OY 360 RPPMAGDTGNGNLQARGAGSSSLGIDAMMAGDAINNMALGKLGAA 403
    ID 09K45 PRELIMINARY; PRT; 365 AA.
DB 339 SAMAGDNGNINLQARGAGSSSLGIDATLTGDAINNMALRLSLAA 382

```

RESULT 2

```

OY 09K45 PRELIMINARY; PRT; 365 AA.
AC 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DE 01-OCT-2000 (Tremblrel. 15, Last annotation update)
GN HRPN.
OS Erwinia herbicola pv. gypsophillae.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
NCBI_TaxID=48984;
RN 1
RA Mor H., Manulis S., Zuck M., Nizan R., Coplin D.L., Barash I.;
RT "Genetic organization of the hrp gene cluster and dsper operon in
    Erwinia herbicola pv. gypsophillae."
    Submitted (May-2000) to the EMBL/GenBank/DBJ databases.
    EMBL: AF211716; AAF/6342.1;
SO SEQUENCE 365 AA; 36784 MW; 6F79B86DD95AC2A CRC64;

```

Query Match 48.2%; Score 1002.5; DB 2; Length 365;
 Best Local Similarity 57.9%; Pred. No. 2.9e-64;
 Matches 223; Conservative 40; Mismatches 83; Indels 39; Gaps 14;

```

OY 1 MSINTSGLASTMOISIGAGNGNGLGTSRONAGLGSALGIGGONNDYVNOAGL 60
    ID 09K45 PRELIMINARY; PRT; 365 AA.
DB 1 MSINTSGLASTMOISIGAGNGNGLGTSRONAGLGSALGIGGONNDYVNOAGL 52
OY 61 TCGMMAMSMNGGGGGLMGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLG 117
    ID 09K45 PRELIMINARY; PRT; 365 AA.
DB 53 TCGMMAMSMNGGGGGLMGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLG 110
OY 118 GNTTSTSTSPDLQALGINSTSONDSTSGTSDTSDS-----DPMQQLKMFSEIMQSL 172
    ID 09K45 PRELIMINARY; PRT; 365 AA.
DB 111 G-----AASSLIDKALDL-----DPTADGGGLSDSNGTYGMSPTIDQLKMFSEIMQSM 157
OY 173 FGDGODGTGSSSGGKOPTGEBQNAKKGVTDALSGIANGSLQILGNGSLGCGGCGAG 232
    ID 09K45 PRELIMINARY; PRT; 365 AA.
DB 158 FGDGODGTGSGKGT-GNQPPTPEDEQNAATKGVTDALTAIVAGGLSOMQSS-GTGGGMMNSIG 214
OY 233 TGLDSSGLGKGLQNLISGVVDYQOQGNVGTGIGKAGIGIOLANDITGTHRSSTRFVNGK 292
    ID 09K45 PRELIMINARY; PRT; 365 AA.
DB 215 L-----GNGGLGKGLQNLISGVVDYQOQGNVGTGIGKAGIGIOLANDITGTHRSSTRFVNGK 271
OY 293 DRAMAEIQQFMDQYPEVFGKPOYOKPGCGEVTGTDKSWAKALSKPDDGMPASMEQFN 352
    ID 09K45 PRELIMINARY; PRT; 365 AA.
DB 272 DRSLAEVGGFMDQYPEVFGKPOYOKPGCGEVTGTDKSWAKALSKPDDGMPASMEQFN 331
OY 353 KAKGMKRPAGDTGNGNLQARGAGSSSLGIDAMMAGDAINNMALGKLGAA 403
    ID 09K45 PRELIMINARY; PRT; 365 AA.
DB 332 KAKGMKRPAGDTGNGNLQARGAGSSSLGIDAMMAGDAINNMALGKLGAA 382

```

RESULT 3
 OY 09K45 PRELIMINARY; PRT; 344 AA.

```

AC 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DE 01-MAR-2001 (Tremblrel. 16, Last annotation update)
GN HRPN.
OS Erwinia chrysanthemi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
NCBI_TaxID=556;
RN 1
RA Vedel R., Berthuis I., Boccara M.;
RT "HrpN of Erwinia chrysanthemi 3937 Role in pathogenicity and
    regulation."
    Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
    EMBL: AJ302027; CAC20124.1;
SO SEQUENCE 344 AA; 34493 MW; 137E535659DCEFO CRC64;

```

Query Match 35.4%; Score 736.5; DB 2; Length 344;
 Best Local Similarity 42.2%; Pred. No. 2.9e-45;
 Matches 171; Conservative 44; Mismatches 111; Indels 79; Gaps 9;

```

OY 13 MOIST-GGAGNGNGLGTSRONAGLGG-----NSALGIGGONNDYVNOAGLITGMMAMM 67
    ID 09K45 PRELIMINARY; PRT; 344 AA.
DB 1 MOIST-GGAGNGNGLGTSRONAGLGG-----NSALGIGGONNDYVNOAGLITGMMAMM 58
OY 68 SMNGGGGGLMGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLG 125
    ID 09K45 PRELIMINARY; PRT; 344 AA.
DB 59 -----GGLSGGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLG 99
OY 126 NSPLDQALGINSTSONDSTSGTSDTSDSDDPMQQLKMFSEIMQSLFG-----DQO 177
    ID 09K45 PRELIMINARY; PRT; 344 AA.
DB 100 Q-----SGSDALS-----KHPDKALDLDLGHDTYTKLTLNQS 130
OY 178 DGTGSSSGGKOPTGEBQNAKKGVTDALSGIANGSLQILGNGSLGCGGCGAGTGLDGG 237
    ID 09K45 PRELIMINARY; PRT; 344 AA.
DB 131 DGTGSSSGGKOPTGEBQNAKKGVTDALSGIANGSLQILGNGSLGCGGCGAGTGLDGG 176
OY 238 SSLGKGLQNLISGVVDYQOQGNVGTGIGKAGIGIOLANDITGTHRSSTRFVNGKDRAMA 297
    ID 09K45 PRELIMINARY; PRT; 344 AA.
DB 177 SSLGKGLQNLISGVVDYQOQGNVGTGIGKAGIGIOLANDITGTHRSSTRFVNGKDRAMA 236
OY 298 KEIGOFMDQYPEVFGKPOYOKPGCGEVTGTDKSWAKALSKPDDGMPASMEQFNKAKGM 357
    ID 09K45 PRELIMINARY; PRT; 344 AA.
DB 237 KEIGOFMDQYPEVFGKPOYOKPGCGEVTGTDKSWAKALSKPDDGMPASMEQFNKAKGM 296
OY 358 IKRPMAGDTGNGNLQARGAGSSSLGIDAMMAGDAINNMALGKLGAA 402
    ID 09K45 PRELIMINARY; PRT; 344 AA.
DB 297 IKRPMAGDTGNGNLQARGAGSSSLGIDAMMAGDAINNMALGKLGAA 341

```

RESULT 4
 OY 09LAW6 PRELIMINARY; PRT; 142 AA.
 AC 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DE 01-OCT-2000 (Tremblrel. 15, Last annotation update)
 GN HRPN.
 OS Erwinia amylovora.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 NCBI_TaxID=552;
 RN 1
 RA STRAIN-EA246;
 SO SEQUENCE FROM N.A.

Query Match 12.0%; Score 249; DB 2; Length 731;
Best Local Similarity 28.7%; Pred. No. 5,6e-10;
Matches 96; Conservative 23; Mismatches 101; Indels 114; Gaps 15

2 SLNTSGTAASTWQISGAGGNNGLGTSRQ-----NAGLGN-SALGIG 45

Dn 402 SAGTGGVAVAS-----GGTGACGATLIGAGHGAGGAGAGNOTGCGVNGNCGAGGAGGAGGAG 455

Qy 46 G--GNODTVNQVLGLTYGMAMMMNMNGGGGLMGGLG-GGLG-----NGLGSGG 93

Dn 456 GOLYNGGDGCGNGAG-----GANINGAGNSDGAAGHAGAGGSARLIGAGHGCGGG 508

Qy 94 LGEGLSNALNDMLGSLWLTLSKGN-----NTSTINSPLD 130

Dn 509 AGGNTAGRRADAIACT-----GGDGNGNGGILLSGNAGAGHGAGGSSSTATTTTGTPT 564

Qy 131 QALGINSTRSQQDDDSGTSDTSDDSPPMQLLKMPSELMSLFEGDPDGTGSSSSGKQP 190

Dn 555 CATGNGNGGNGCAGGTAGFTGSG-----GIGNGGAGGTGNNG----- 602

Qy 191 TEGEONAYKKGYTDAL-----SGLMGNGLSOLLNGGLG--GGGNGAGTGLDSSLSGK 243

Dn 603 -----VALSVGSTGLGNGSGGGLGGGGGSLFNGGAGAGGVATGNGSGSIGPASPVG 657

Qy 244 G-----LQNLSEPVYDQLGNNAVGTGIMKAG 270

Dn 658 GKGGVGAAGLAG-----QIQNGSGSGSGGAG 686

RESULT 8

ID 053552 PRELIMINARY; PRT, 1381 AA.

AC 053552;

DT 01-JUN-1998 (TREMBLrel, 06, Created)

DT 01-JUN-1998 (TREMBLrel, 06, Last sequence update)

DT 01-OCT-2000 (TREMBLrel, 15, Last annotation update)

DE PGFS-FAMILY PROTEIN.

GN RV3507 OR MTW023.14.

OS Mycobacterium tuberculosis.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

CC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.

OX NCBI_TaxID=11773;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-H37RV;

RX MEDLINE=98295987; PubMed=9634230;

RA Cole S.T., Brosch R., Parkhill J.J., Garnier T., Churcher C., Harris D., Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F., Badcock K., Basham D., Brown D., Chillingworth T., Connor R., Davies R., Deakin K., Felwell T., Gentles S., Hamlin N., Holroyd S., Hornsby T., Javelis K., Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;

RA "Taylor K., Whitehead S., Barrett B.G.;"

RT *Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.";

RL Nature 393:537-544(1998).

EMBL AL022022; CAAL7744.1; -

DR HSSP; P00778; 20UL.

DR TuberculList; RV3507; -

DR InterPro; IPR000084; -

DR InterPro; IPR002173; -

DR Pfam; PF00934; PE.1.

DR ProDom; PD001223; -1.

DR PROSITE; PS00583; PRFB_KINASES_1; UNKNOWN_2.

SQ SEQUENCE 1381 AA; 110624 MW; CA09676BD07F6482 CRC64;

Query Match 11.9%; Score 248; DB 2; Length 1381;

Best Local Similarity 26.8%; Pred. No.1.5e-09;

Matches 117; Conservative 24; Mismatches 180; Indels 116; Gaps 16;

Qy 7 GIGASTMQTISIGAGGNNCLIGTRONAGIAGNSALGIGGQNODTVNQVLGLTYGMAMM 66

Dn 822 GIGAGCTTTVTGNGCMAG-DGGNGSMAGAGNGSGDGRGMT----- 864

Qy 67 MSMGGGGLMGGLGCGLCNGJGSGGGLGEGLSNALNDMLGSLWLTLSKGNNTTSTTN 126

Db 865 -SCASSGSGNGNAGTACSGAGGTGCTGLSGGNGNGNGN-----GGGCGNCAHGTVC 919
QY 127 SPLDQALGINSTONDDSTGDTSDSDP--MOQLIKFSEIMQSLFPGDQDGTGSS 184
Db 920 AOFVPAVSLPTPNCAGANGNGTGSNGAPGAPGPTTGGNAGSOGIGCGGNGGGK 979
QY 185 SGGR-----OPTBEQNAKKGVTDALSLGMLNGLSOLLNGGIG-----224
Db 980 GGDADAVNVVFMPTGPQ-----AATGTAGSAGDPTGCGGCGTGPSPVAVAPP 1029
QY 225 -----GGGCGNAGTGL-----DSSSLGKGLN-----LSGPVDQQLGNVGTGIM 267
Db 1030 TPTTQVQGGDGGGAGGAGGTSTANDGTATNGKRGCGVGSILGPP-----GGNGGTG--- 1080
QY 268 KAGIQALNDIGTHHSTSTSFVNKDRAMAKELGOFMDQYPEVEGPKQYKPGCEVKT 327
Db 1081 -----GNASATGTNGVANAGNGKGGDGGQF-----GAGGNGAGGSV--T 1119
QY 328 DKSMAKALSKRDDGKMPASMEQFNKAKMKRPMAGDPTGNGTQARAGSSILGIDAM 387
Db 1120 DGSAGSTAGNGNG-----NATNGTI---AGQPAGNGSAGKGGDGNIAAGA 1166
QY 388 AGDAIN--NNALGKIGA 402
Db 1167 TGTAGNGNGNGNDGA 1183

RESULT 9
ID 050396 PRELIMINARY; PRT: 588 AA.
AC 050396;
DT 01-JUN-1998 (TREMblrel. 06, Created)
DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)
DE 01-OCT-2000 (TREMblrel. 15, Last annotation update)
DE PGFS-FAMILY PROTEIN.
GN RV3367 OR MTY004.25.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE-98295987; PubMed-9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F., Badcock K., Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K., Felkell T., Gentles S., Hamlin N., Holroyd S., Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy J., Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the RT complete genome sequence."
RL Nature 393:537-544(1998).
RL EMBL: AL009198; CAAL5752.1;
DR EMBL; AL009198; CAAL5752.1;
DR InterPro: IPR000084;
DR InterPro: IPR002173;
DR Pfam: PF00934; PE; 1.
DR PRODOM: PD001223; -; 1.
DR PROSITE: PS00583; PKK_KINASES_1; UNKNOWN_1.
SQ SEQUENCE 588 AA; 49708 MW; 067B84097F61DAF1 CRC64;

Query Match 11.9%; Score 247; DB 2; Length 588;
Best Local Similarity 30.4%; Pred. No. 6e-10;
Matches 117; Conservative 16; Mismatches 154; Indels 98; Gaps 21;

QY 18 GGAGG--NNGGLGTSRONAGTGCNSALGIGGNGNDTVNQLAGLLTGMMMMNMGGGGL 75
Db 283 GGAGGDANGNPANTSIANNAGGNGAAG-GDGGANGGAGGAGQAASAGSVGCGDGNCG 341

QY 76 MGG-----GLGGGLGNGLAGSGGLGEGLSNALNDMLGSLNTLGSKGNNTSTNSPLDQ 131
Db 342 AGGTCTGNHAGAG-GAGGAGGAGRGWLVG--NGNGCN-----GAAGGGAIGTGC--- 390
QY 132 ALGINSTONDDSTGDTSDSDPQQLIKFSEIMQSLFPGDQDGTGSSSGKOPT 191
Db 391 AGGV-PANQGNLSALT-----OPVGGDGGCGNGGTG---T 424
QY 192 EGEONAYKKGYTDALSG-IMGNGLSOLLNGGIG--GGGNGAGTGLDSSSLGKGLQNL 248
Db 425 GGRGDDGSGGAGGASGMLMNGNG--GNGGTGSGGVGNGGIGDAG--GNAATST 481
QY 249 SGPPYDQQLGNAVGTGIMKAGIQALNDIGTHHSTSTSFVNKDRAMAKELGPMQY 308
Db 482 SIPEFANG-GNG--GAGGDAG-----HGTGGDGGDGGHAGTGRNGGL----- 522
QY 309 EYFGKPYQKPGQEVKTRDDKSMAKALSKPDDGWTMPASMEQFNKAKMKRPMAGDPTGN 368
Db 523 -----AGCHANGNGGGGTGAGGTHG--TPGSGN-----AGTGT 557
QY 369 GNTQARGAGSSSLGIDAMAGDAIN 393
Db 558 GNAIDSTNGCPGSDG---LGDAFN 578

RESULT 10
ID 053557 PRELIMINARY; PRT: 1079 AA.
AC 053557;
DT 01-JUN-1998 (TREMblrel. 06, Created)
DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE PGFS-FAMILY PROTEIN (FRAGMENT).
GN RV3512 OR MTY023.19.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE-98295987; PubMed-9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F., Badcock K., Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K., Felkell T., Gentles S., Hamlin N., Holroyd S., Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy J., Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the RT complete genome sequence."
RL Nature 393:537-544(1998).
RL EMBL: AL022022; CAAL7749.1;
DR EMBL; AL022022; CAAL7749.1;
DR InterPro: IPR002202;
DR PROSITE: PS00318; HMG_COA_REDUCTASE_2; UNKNOWN_1.
FT NON_TER 1
SQ SEQUENCE 1079 AA; 81163 MW; A79718CDB74B97D CRC64;

Query Match 11.9%; Score 246.5; DB 2; Length 1079;
Best Local Similarity 26.2%; Pred. No. 1.4e-09;
Matches 118; Conservative 26; Mismatches 164; Indels 143; Gaps 16;

QY 5 TSGLGASTMQISTGAGGNGGLGTSRONAGTGCNSALGIGGNGNDTVNQLAGLLTGMM 64
Db 636 TGGAGNGGCGGANGGAGGAGSGGTGNGNGAGGADGADGNGNGTGNCGNG----- 689
QY 65 MMSMMNMGGGGL--MGGGLGGGLGNGLAGSGGGLGEGLSNALNDMLGSLNTLGSKGN 119
Db 690 -----GNGGIAGMGNGAGGTGSGNGNGSGS--GNGNGMGNGNSGTGSGDGGAGGN 740


```

QY 120 NTSTNSPLDQALGINST-----SONDSTGCTSTSDSPMOQLMFSEIM 169
Db 741 GGAAGTGTGGTGG-GLTGTGTGGTGGTGGGNGGADNTAN-----MT 785
QY 170 QSLFGDQDGTGSSG-----GKOPTGEONATKKG-----VTDLASG 208
Db 786 AAGAGDGGNGGDDGGFGGAGAGGGLTGANCTGGCAGAGCGNGAIGCHPLTDDPGG 845
QY 209 LMGSLQLGNGSLG-----GGGNGMGTGIDGSSISGKGLQNLSCP 251
Db 846 NGGTG-----GNGTGTGTGAGAGISLGGTGGGCGNGNGTGGEGEVGAG-----893
QY 252 VDYQQLGNVGTGIGMAGIQLANDIGTHRSSTRSFVNKGRAMAKEIGQFMQYEVF 311
Db 894 -----GTGGAAGNGGDDGTGGTGGGAGGTGTGTGGLD-----PRVG 935
QY 312 GKPYQAGPQEVYTDKSNAKALSKPDDGMPASMEQFNKAKGMIKRPAGCTGNGNL 371
Db 936 G-----SGDDGTGGSGGAGNGGNG-----GNAGAGGNGNG 967
QY 372 QARGAGSSIGIDAMMAGDAINMALKLGAA 402
Db 968 GTGAGG--TGGTGGNGGDAEPGVPPAGGA 996

RESULT 11
OS3395 PRELIMINARY: PRT; 1538 AA.
AC 053395;
DT 01-JUN-1998 (TREMblrel. 06, Created)
DT 01-AUG-1998 (TREMblrel. 07, last sequence update)
DT 01-OCT-2000 (TREMblrel. 15, last annotation update)
DE PGFS-FAMILY PROTEIN.
PE PGFS OR RV3345C OR MTW004.01C-MTW016.45C.
GN Mycobacterium tuberculosis.
OS Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
[1]
SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RA MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagsels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [12]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RA Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RL EMBL: AL021841; CAI17117.1;
RA HSSP: P00441; ISOS.
DR Tuberculin; RV3345C;
DR InterPro: IPR000084;
DR InterPro: IPR002173;
DR Pfam: PF00934; PE: 1.
DR ProDom: PD001223; -.
DR PROSITE: PS00583; PFKB_KINASES_1; UNKNOWN_3.
SQ SEQUENCE 1538 AA; 129386 MW; 788F0B2095587592 CRC64;

```

```

QY 7 GLASTWQISIGAGGNGNGLGTSTRONAGLGG-----NSALGLGGGNDPVTN-----54
Db 249 GTG-----GIGGNGGAGLFG-----NGAGAGAGALPGAGLNGSDGSDGNGTGCN 298
QY 55 -QLAGLLTGMAMMAMMMGGGLMG-----GGLGGGLGN-GLGSSGG 93
Db 299 GGRGGLLVG-----NGAGAGAGAGVGGDGGKGGADPSPFAVNNAGAGNGHGNGPVGAGG 354
QY 94 LGEISNANLMDLGSILNTLGSKGGNNTTSTNPLDQALGINSTSONDSTGTDSTSD 153
Db 355 AG-GLLAGAHMAAGATPTSGNGGDDGIGATVANSPL-QAGAGAGNG-CHGLVNGGTGG 411
QY 154 SSDPMQQLKMFSEIMQSLFGDGTGSSSGGKOPTGEONATKKGVTDALSLGNG 213
Db 412 AG-----GAGHAGTGTATGLTLOPTGGNGT-----NGAGAGGNGNGNG 450
QY 214 LSQLLNGSLG-----GGGNGMAGTGLDSSISGKGLQNLGSPVYQQLGNVGTGIGMK 268
Db 451 GAQ-HGDOGVGKGGAGSGGAGNGGFDATLGSFADGGNG-----GNGGKGGDGGK 502
QY 269 AGIQLANDIGTHRSSTRSFVNKGRAMAKEIGQFMQYEVFSGKPYQAGPQEVYTD 328
Db 503 AG-----DGGAGAGDVTLAVNQGAG-----DGGNGGEVGVGG 536
QY 329 KSWAKALS-----KPDGDMTPASMEQFNKAKGMIKRPAGDGN-----GN 370
Db 537 KGGAGVSAHPALNGSAGANCTAPTSGGNGGAGATPTVAGENGAGNGHGSGVGN 596
QY 371 LQARGAGSSL--GIDAMMAGDAINMALKLGAA 403
Db 597 GGAGAGAGNGVAGTGL-ALNGNGNGNGIGGNGGSA 631

RESULT 12
OS33215 PRELIMINARY: PRT; 1660 AA.
AC 0533215;
DT 01-JUN-1998 (TREMblrel. 06, Created)
DT 01-JUN-1998 (TREMblrel. 06, last sequence update)
DT 01-OCT-2000 (TREMblrel. 15, last annotation update)
DE PGFS-FAMILY.
GN RV2490C OR MTW008.46C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
[1]
SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RA MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagsels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [12]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RA Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RL EMBL: AL021246; CAI16067.1;
RA Tuberculin; RV2490C;
DR InterPro: IPR000084;
DR InterPro: IPR000228;
DR InterPro: IPR002173;
DR Pfam: PF00934; PE: 1.
DR ProDom: PD001223; -.
DR PROSITE: PS00583; PFKB_KINASES_1; UNKNOWN_1.
DR PROSITE: PS01287; RTC; UNKNOWN_1.
SQ SEQUENCE 1660 AA; 133124 MW; 3A89CE12C0FA945 CRC64;

```

Query Match 11.88; Score 246; DB 2; Length 1538;
 Best Local Similarity 27.28; Pred. No. 2.3e-09;
 Matches 124; Conservative 32; Mismatches 168; Indels 132; Gaps 21;


```

Db      521  GGGAGSVGGGGGAGGGDGFVGA-----GGNGAGNAGTGVGVNANGN-----CGSA 568

Qy      61  TGGMMMSHMMGGGGGLAGGGLAGGLGAGLGGSGGLGIBELSNALN--DMLGSLNT-LGSKG 117
      11  :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
      569  TG---ALAAVAGAGAAAGDNTSGTG--GFGGAGGSARGLFALFGAGAGAAAGDASTGVGGPG 624
Db

Qy      118  GNNNTSTTNSPLDQALGINST--SONDSTGCTDSTSDSPMOQLKMFSEIMQSLFGDG 176
      11  :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
      635  GPGGTGTAASPFGALAIIGAGAGAGGAGTSCATGACAGDGVFEGT-----AVLGLRG-G 678
Db

Qy      177  QDGTGSSSSGGKQPTFEGEONAYKKGVTDALSGT--MNGLSLQLG-----NGGLGGGGGN 230
      11  :::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
      679  AAGAGGAATG-----DCAATGAGAGGFGGAGAGANLAFGLGFVYLHGAGAGAGTA 725
Db

Qy      231  AGTGLDSSSLGKGLQNLGSPVYDQQLGNMVGTCIGKKAG-----IQALNDIGTHRHSST 285
      11  :::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
      726  TGTGTGNGGAGGGGGG---LSPV-----ILGIGAGGDDGGGALCVLGMGW----- 769
Db

Qy      286  RSYFNKQGRAAAKELIGCFPMQYPEVFGKPKYQKPFQCFEYKTDKDSMAKALSKPDDDGMP 345
      11  :::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
      770  ---DGGGGCAVAVGLAVG-----GAG-----788
Db

Qy      346  ASMEQFNNAKEMIKRPMAGDPTGNGNLQARGAGSSSLCIDMMAGDANINNALCKLTG 401
      11  :::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
      789  ---GAGG---AAPTGG---GAGGN-----GGDALGLVLGVGGNG 818
Db

```

RESULT 15

ID	053556	PRELIMINARY;	PRT;	714 AA.
AC	053556;			
DT	01-JUN-1998 (TREMBLrel. 06, Created)			
DT	01-JUN-1998 (TREMBLrel. 06, Last sequence update)			
DT	01-OCT-2000 (TREMBLrel. 15, Last annotation update)			
DE	PG8S-FAMILY PROTEIN.			
GN	RV3511 OR MTW023.18.			
OS	Mycobacterium tuberculosis.			
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;			
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.			
OX	NCBI_TaxID=1773;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-H37Rv;			
FX	MEDLINE=98255987; PubMed=9634220;			
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,			
RA	Gordon S.V., Eiglmeier K., Gass S., Barry C.E. III, Tekala F.,			
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,			
RA	Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holroyd S.,			
RA	Honnsby T., Jagels K., Krogh A., McLean A., Moule S., Murphy L.,			
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,			
RA	Raynor K., Whitehead S., Barrell B.G.;			
RT	*Deciphering the biology of Mycobacterium tuberculosis from the			
RT	complete genome sequence. "			
RL	Nature 393:537-544(1998).			
DR	EMBL; AL022022; CAA17748.1; -			
DR	TubercuList; RV3511; -			
DR	InterPro; IPR000084; -			
DR	InterPro; IPR002173; -			
DR	Pfam; PF00934; PE: 1.			
DR	ProDom; PD001223; -; 1.			
DR	PROSITE; PS00583; PFKB_KINASES.1; UNKNOWN.1.			
SO	SEQUENCE 714 AA; 59990 MW; E66D2A7FAF5CF0E1 CRC64;			

Query Match	11.4%;	Score 236.5;	DB 2;	Length 714;
Best Local Similarity	28.0%;	Pred. NO. 4.3e-09;		
Matches 120;	Conservative 27;	Mismatches 161;	Indels 121;	Gaps 22

QY 5 TSGLCASTMVISIGGAGCNGNC-LIGTSKONAGLGCNSALGLGGGNDNTVQIAGLLTGM 63
| : | | | | | | | | | : | : | : | |
Db 215 TGGIG-----GGPFGNGGMLTG-----NGHGCG-AGSIGGSSG-----AGGNGCW 254

```

Oy 64 MMMSMMGGGGLMGGLGGGLGN-----GLGSGGLEG-----LSNLND 104
      ::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 255 LLNGNGIGGAGGTGGAGGTGGMAAMLGGGTGGAGTGGNGHGGNGWLLGNGNG 314
Oy 105 MLGSGLN--TLGSKGNNNTTSTNSP--LDQALINSTSNDSTGTSTSDSPMQ 159
      |||::|||::|||::|||::|||::|||::|||::|||
Db 315 GLGGGDDGTGGGHGNG-----GNPGLLCTAGGGGNGAGSTGTAGGSGGNG 364
Oy 160 QLKKEFSEIQLSLEGGDGC-----TQSSSGGKQPTBEGDNL-YKKGYDALSGLMNG 213
      |||::|||::|||::|||::|||::|||::|||::|||
Db 365 -----GDGTGGRGGLLMGAGGHHGTTGGAGAGCAVNGGAGGAGGAGNG 410
Oy 214 -----LSLLNGGLG-----GGGGGNAGTGLDG--SLGKGKQLNLSGPVYQQLNAV 261
      ::|||::|||::|||::|||::|||::|||::|||
Db 411 GAGGCAALLFGRGGTGGAGGTGGDGGGGGDDFTGMALGGTGGSGGTGD-----GAP 465
Oy 262 GTGTCMKAGIALNDICTHRHSSTRSPFNKKDRAMAKEITGPFMOQYEVFGKPOYQKPG 321
      |||::|||::|||::|||::|||::|||::|||
Db 466 GNGGAGGAG-QLSHISGYVAGASGKAGAGCTGGNGGAGSAGADAPAGSGAMGSTGFAGAG 524
Oy 322 QEVKTDKSMKALKSPDDGCMTPASMEQFNKAKGMITKRPAGDTGNGNLQAR-GAGGSS 380
      ::|||::|||::|||::|||::|||::|||
Db 525 -----GDGG-----NGGSGASQCGNGCGNGCGTGGKGTGGAG 558
Oy 381 L-GIDAMMA 388
      ::|||::|||::|||::|||::|||
Db 559 MNSLDPLIA 567

```

```
Search completed: October 25, 2001, 11:47:11
Job time: 192 sec
```